AN APPROACH TO FREQUENCY ANALYSIS USING EXPERT SYSTEM TECHNIQUE

A Thesis Submitted
in Partial Fulfilment of the Requirements
for the Degree of
MASTER OF TECHNOLOGY

by

OM PRAKASH MISHRA

to the

DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

November, 1990

To

Babuji and Maa

CERTIFICATE

This is to certify that the thesis titled **"AN APPROACH TO FREQUENCY ANALYSIS USING EXPERT SYSTEM TECHNIQUE"** submitted by Shri Om Prakash Mishra, in partial fulfilment of the requirements for the degree of Master of Technology of the Indian Institute of Technology, Kanpur, is a record of bonafide research work carried out by him under our supervision and guidance. The work embodied in this thesis has not been submitted elsewhere for a degree.

November, 1990

(B.Datta)
Asst.Professor

(S.Ramaseshan)
Professor

Department of Civil Engineering
Indian Institute of Technology
Kanpur - 208016

ACKNOWLEDGEMENT

I express my deep sense of gratitude and indebtedness to my guides Dr.S.Ramaseshan and Dr.B.Datta for their invaluable guidance and constant encouragement for the successful completion of this work. It was a very exciting experience to have worked with them.

I am grateful to all my teachers who taught me at I.I.T. Kanpur.

I would like to acknowledge Mr.JayKumar for his brotherly affection and constant help during my stay at I.I.T. Kanpur, both academically and otherwise.

Special thanks are due to Mr.Reddy, R.E. of ACSE for his kind co-operation and guidance in proper understanding of the software used.

Thanks are also due to Ajay, Ammu, Babli, B.Tirupati, Pitam, Pradeep, Ravi, Satish, and Shok Kesri for their direct and indirect help in this endeavour.

Finally I am extremely grateful to my parents for their constant inspiration, blessings, and expedient advice in this endeavour.

Om Prakash Mishra

C O N T E N T S

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Population mean

Population standard deviation

Coefficient of skewness for population

Coefficient of kurtosis

Coefficient of variation

Length of data

ABSTRACT

Frequency analysis of hydrologic data is a knowledge domain where lot of decisions are to be taken based on human intuition, experience, and subjective judgement. Being a data dependent technique the reliability of frequency analysis is always questionable. To fit a theoretical distribution function to a set of observations and to draw conclusions therefrom, require a combination of routine numerical/statistical computation and human expertise in the form of heuristics. In order to address these problems, a program, Frequency Analysis of Continuous Hydrologic Variables with an embedded Expert System (FACHVES) has been developed. This program was developed in FORTRAN for microVAX-VMS environment. It consists of methods for computerised statistical analysis supported by an expert decison support system. The framework consists of four parts: main program, subroutines, expert system interface, and a knowledge base. The developed work completed so far is preliminary in nature and it provides the framework for future development of a more comprehensive computer program package.

The developed package for FACHVES was tested with two different data sets representing recorded streamflow data at two different sites in India. The results of this testing shows the proper functioning of the package at this stage of development. These results also established the feasibility of this approach for statistical analysis of the hydrologic data by a combination of computational procedures and expert advice.

CHAPTER 1

INTRODUCTION

1.1. General

With the development of digital computers, engineering design and analysis got new dimensions. Fast and efficient computing due to computers made several complex analysis techniques, e.g., numerical analysis, statistical analysis etc., computationally feasible. Higher level languages like FORTRAN, BASIC, C and PASCAL were developed and a number of programs useful for engineering problem solving have been written in these languages. Though initially these programs were useful to solve a problem, they were not able to solve complex problems involving a number of ordinary and partial differential equations, logical constraints, or analyse problems involving uncertainties. Therefore for the processes which are not well understood, involving uncertainties due to lack of data in quality and quantity, simulation analysis had to be resorted to. Later on it was realised that in several cases a number of complex problems depended upon intuition and subjective judgment of the analyst and could not be solved by routine analytic techniques. Development of languages like LISP and PROLOG in which logical approaches were easier to be programmed, facilitated the design of a system to sort out these problems. Initially the intention behind these approaches was to design such a computing system which can think and derive conclusions from logical and

numerical analysis just as a human being. After a number of attempts the researchers in this field realised the impossibility of such a universally wise system. The recent trend of research in this field is to design limited systems for specific domains.

Water resources engineering (WRE) deals with many processes whose physics is not well understood and many a time decisions are to be taken on the basis of past experience, intuition etc. Moreover data available for analysis are also subject to many observational or computational errors. Therefore problem solving techniques in this field of engineering design also need such a system which can incorporate the domain specific expertise with computational techniques.

Thus computer aided analysis and design in engineering seems to be presently (1990) undergoing a dramatic change in that the numerical, analytic and logical capabilities of a digital computer are extended with the heuristic capabilities of artificial intelligence and expert systems in efficiently solving problems in the areas of analysis and design.

1.2 Water Resources and Expert System

1.2.1 General

Intelligence in human nature is associated with one or more of the following abilities:

*. to respond to situations very flexibly

*. to make sense out of ambiguous or contradictory messages

*. to recognise the relative importance of different elements of a situation

*. to find similarities between situations despite differences which may separate them; and

* to draw distinction between situations despite similarities which may link them.

Above abilities are nothing but common sense of people and they come naturally to them by experience, knowledge of heuristics and largely due to mental faculties. **Artificial Intelligence (AI)** is simply a way to impart these abilities, i.e., intelligence to a computer (Levine, 1988). This can be accomplished by studying how people think when they are trying to make decisions and solve problems; breaking those thought processes down into basic steps; and then designing a computer program that solves problems using those same steps. Fig. 1.1 schematically shows these steps. Thus AI can be defined as that part of computer science concerned with designing intelligent computer systems, i.e., systems that exhibit the intelligent characteristics of human beings like understanding language, learning, reasoning, solving problems and so on.

The field AI encompasses robotics, game playing, the automated translation of language and a major subdiscipline called **Expert System (ES)**. ES is concerned with the development of computer software that can partially represent human knowledge and utilise that knowledge to solve complex problems within a specific domain (Johnston, 1985).

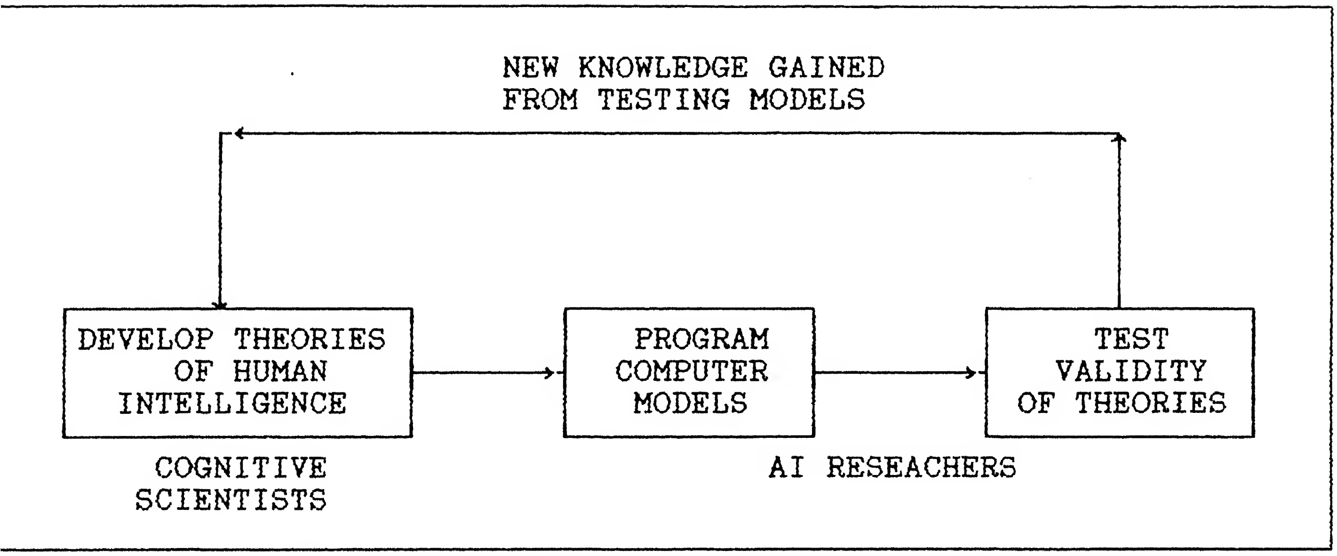

FIG. 1.1 : "The Feedback Loop" in AI Computer Modelling
(Mishcoff, 1986)

1.2.2 Need of ES in water resources engineering

Water resources engineering deals with planned development and management of water, the spatially and stochastically varying natural resource above and below ground. Optimal utilisation of water requires the evaluation of the occurrence, distribution and variability in time and space of the water resources as well as their extremes like flood and droughts. In all the four stages of designing a water resource system, viz. planning, development, management and design, a large amount of uncertainties are involved. These uncertainties can arise due to hydrology, hierarchical and multistage nature of decision making, financial and economic variabilities, and implementation techniques, social constraints, and changing national conditions. An efficient approach to the design of a water resource system should be able to take care of these uncertainties as far as it can. A large amount of domain dependent expertise is needed in planning, design, construction and integrated operation of water resource systems and they are often heuristic in nature. Therefore, water resources engineering has a potential and justification for ES applications.

1.2.3 ES applications in water resources engineering

For many areas of water resources ES are being developed and applied. They include areas like database management, information systems, and water quality management (Smoilier, 1985; Datta and Peralta, 1986; Arnold et al, 1989; Simonovic, 1989; Datta etal, 1990;); selection of design data as design

storm (Nielson, 1986) or flood estimation (Fayegh and Russel, 1986) and appropriate treatment technology for water supply/sewage (Arnold, 1986). ES has been applied also for data analysis (Wilson, 1986), hydrologic modeling and parameter estimation (Engman, et al, 1986; Delleur, 1988), tank irrigation system (Oswald, 1989), choice of model to be used and preparation of input data for a reservoir system (Savic and Simonovic, 1989) and for an urban storm sewer system (Lindberg and Nielson, 1986). Also, management of multipurpose system of reservoirs integrated with conjunctive use of surface water and ground water are complex problems which need human expertise, common sense and heuristics. ES like SERPES for sewage rehabilitations planning process and WADNES for handling emergencies in a water supply network (Ahmad et al, 1989) have been successfully implemented.

Datta and Peralta (1986) proposed a methodology for embedding a statistical pattern recognition system within an expert decision support system for identifying unknown sources of ground water pollution. This method was developed and tested for application (Datta et al, 1990). The concept of combining the decision making capabilities of an expert decision support system and a geographical information system (intelligence GIS) was proposed in Arnold et al, (1989). The primary purpose of such a system is to incorporate heuristic knowledge regarding measurement and other uncertainties in the estimation of hydrologic or water quality variables.

These applications clearly show the possibilities of ES application to water resource engineering.

1.3 Objective of the Study

Water resources engineering encompasses a vast domain. Since it is not possible to consider in this study each and every aspect, a particular area, viz., only probabilistic approach has been considered and frequency analysis of hydrologic data is chosen. Frequency analysis of hydrologic data is a necessary first step in design of hydraulic structures, control of extreme hydrologic events, management of water resources etc. This field of water resource engineering was found interesting and suitable for ES applications as it deals with data lacking in quality and quantity, having observational and computational errors and sparseness. Moreover, there is no universally acceptable methodology for fitting a probability distribution to a given set of data and if at all, only a limited knowledge about the parent distribution of samples may be available. Physical processes resulting in a high or low precipitation or streamflow are also not well understood. Often the presence of outliers, measurement errors or occurrence of rare combination of physical processes or multiple distributions makes the problem more complicated. Hence an efficient approach to frequency analysis requires incorporation of statistical inference tools as well as heuristic knowledge of knowledgeable human experts in this field. Moreover, ES application to frequency analysis seems to be a virgin area of study. The major objectives of the study are as follows:

1. To develop a framework for an ES for frequency analysis of hydrologic data,

2. To implement an interactive FORTRAN based program with an ES in IIT Kanpur computing environment; and
3. To develop and test an interactive program for frequency analysis of hydrologic data.

1.4 Scope of the Study

A variety of computer programs in Fortran language are available in IIT Kanpur for fitting one or more probability distribution. Furthermore an HP-9000 computer system with a unix operating system, a MicroVAX II computer system with a VMS-VAX operating system and IBM PCs, are available for the study. A number of ES shells of different capabilities are also available. However the scope of study was limited because of time constraint and has been limited to:

1. Fortran based program with an ES tool (CLIPS, as identified in Sec. 2.4.1)
2. Frequency analysis of continuous hydrologic variables and testing only a few sets of data
3. A limited number of methods and approaches to frequency analysis, and
4. MicroVAX-VMS environment.

1.5 Organisation of the Study

The study is reported in the following sequence:

1. Introduction to expert system; need and application in water resources engineering; objectives, scope, and organisation of the study (Chapter 1).

2. A brief introduction to expert systems, building expert systems, different expert system tools, and selection of ES tool (CLIPS) for the study (Chapter 2).
3. Introduction to CLIPS and some of its salient features (Chapter 3).
4. Problems involved in frequency analysis in hydrology; expertise available in frequency analysis; introduction and flow charts of the FACHVES (Frequency Analysis for Continuous Hydrologic Variables with Embedded Expert System) program, application, and discussion of results (Chapter 4) and
5. Summary, conclusion and suggestions for future study (Chapter 5).

CHAPTER II

EXPERT SYSTEMS

2.1 Introduction

Human experts in any field are frequently in great demand and are also generally in short supply. AI presents a solution to this problem through an **expert system (ES)** which is a computing system capable of representing and reasoning about some knowledge-rich domain with a view to solving problems and giving advice (Jackson, 1986). It is also known as **knowledge - based expert system (KBES)**.

Gasching (1981) defines KBES as "interactive computer program incorporating judgment, experience, rules of thumb, intuition and other expertise to provide knowledgeable advice about a variety of tasks". All the knowledge in ES is provided by people who are experts in that domain and it has got expert information essentially for the given domain. So ES's act as intelligent assistants to human expert, and also provide assistance to people who otherwise might not have access to expert advice.

2.2 Difference Between KBES and Other Programs

Although both KBES and **database** programs feature the retrieval of stored information, the two types of programs differ greatly.

A database program retrieves facts that are stored, while an KBES uses reasoning to draw conclusions from stored facts (Fig. 2.1). Interactive programs that incorporate graphics and some sort of expertise in the form of constraints and limitations, assumptions and approximations can be considered to give results as advices rather than answers. Yet KBES differs from traditional computer programs in that (Adely, 1985):

1. Expert systems are knowledge intensive programs;
2. In a rule based expert system, expert knowledge is usually divided into many separate rules;
3. The rules forming a knowledge base or expert knowledge is separated from the methods for applying the knowledge to the current problem. These methods are referred to as inference mechanism or rule interpreter;
4. Expert systems are highly interactive;
5. Expert systems have user friendly intelligent user interface; and
6. Expert systems, to some extent, mimic the decision making and reasoning process of human experts. They can provide advice, answer questions and justify their conclusions.

Nonintelligent programs generally follow a well defined algorithm that specifies explicitly how to find the output variables for any given input variables. It is called **procedural programming**. On the other hand, in the intelligent program used in KBES the behavior of the program is not explicitly described by the algorithm. The sequence of steps followed by the program is influenced by the particular problem presented to it. This is

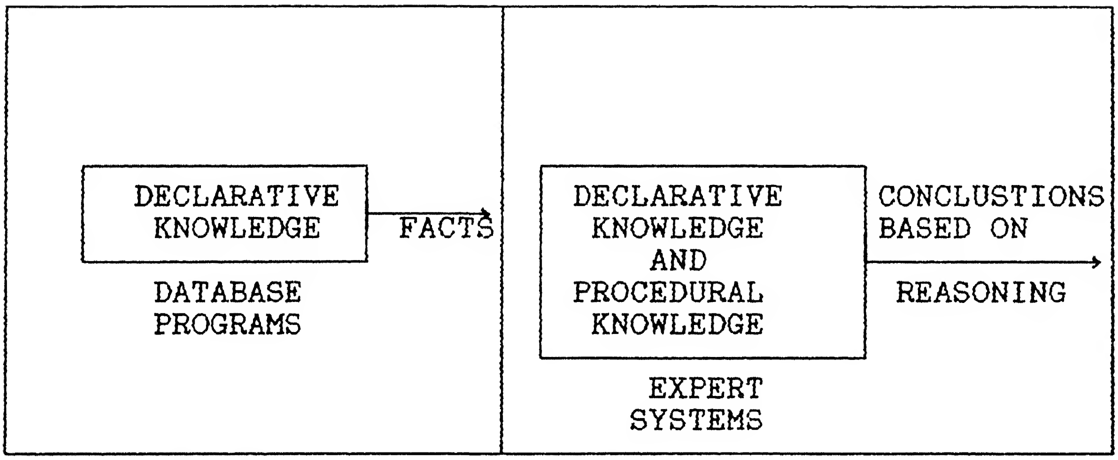

FIG. 2.1 DATABASE PROGRAM VS. ES
(Harmon and King, 1985)

called **declarative programming**. Declarative programming is more efficient than procedural programming in that it enables computer to absorb new information at intermediate steps and further action can be referred to without unnecessarily going all the way back to the beginning of the program or disturbing the already established other facts (Levine, 1988).

2.3 Building Expert Systems

2.3.1 Architecture of KBES

A KBES generally has four principal components (Fig. 2.2). They are, a knowledge base, working memory, and inference engine and a user interface. As KBES vary in design, they may have other components also, e.g., graphics, system analysis and other software.

a. Knowledge base. A knowledge base contains both declarative knowledge (facts about objects, events and situations) and procedural knowledge (information about courses of action) which may be scientific, analytic or heuristic rules (Fig. 2.3). Although many knowledge representation techniques have been used in ES, the most prevalent form of knowledge representation currently used in ES is the **rule based production system** approach. The rules have generally two parts, conditions and actions. The rules are fired when the conditions are matched with the facts. The actions can be for processing instructions or control instructions. The rules may include metarules which are rules about rules.

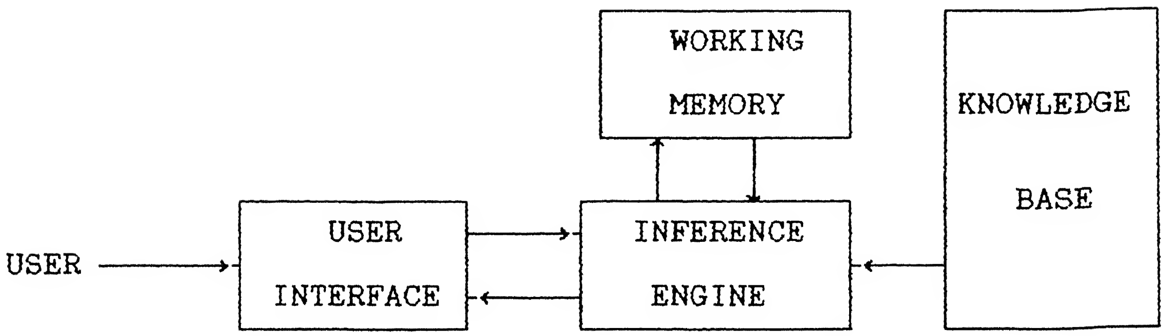

FIG. 2.2 ARCHITECTURE OF A TYPICAL ES

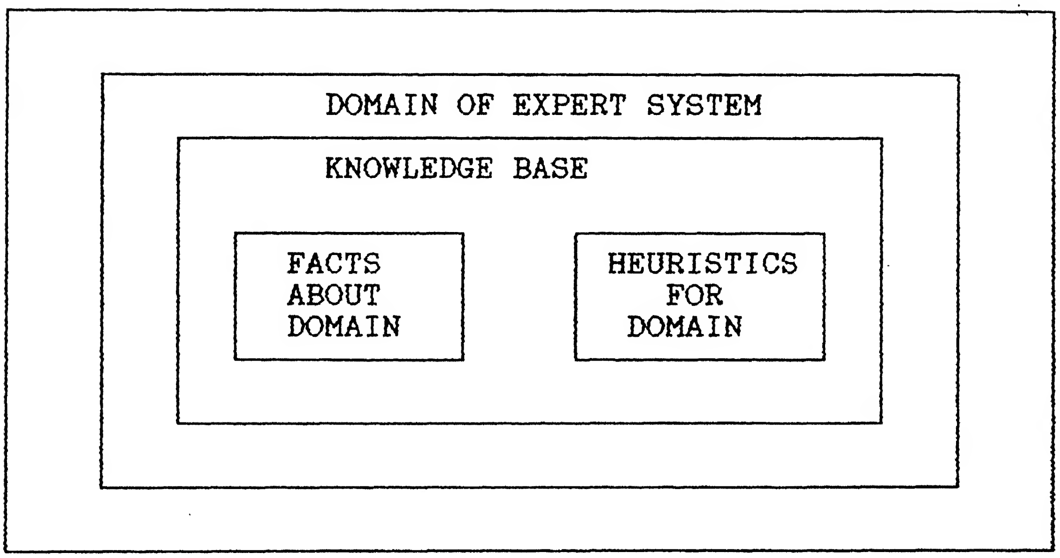

FIG. 2.3 : THE COMPONENTS OF THE KNOWLEDGE BASE OF AN ES (Schank and Childers, 1984)

b. Working memory. It is the current active set of the knowledge base and may include a knowledge management module.

c. Inference engine. It is the most crucial component of ES since it matches and manipulates the database for problem solving. It is the inference mechanism which also provides justification for advice from ES. Three formal approaches used in this case are production rules, structured objects and predicate logic. Production rule consists of a rule set, a rule interpreter which specifies when and how to apply the rules and working memory that holds data, goals or intermediate results. Structured objects use vector representation of essential and accidental properties. Predicate logic uses prepositional and predicate calculus.

d. User interface. It is the communication module which provides bi-directional exchange of information between user and system.

2.3.2 ES Techniques

The order of execution of the rules and/or procedures in an ES is governed by the inference engine in terms of the problem solving strategy used. Maher (1986) considers two approaches:

a. The derivation approach: It involves deriving a solution that is most appropriate for the problem from a list of predefined solutions stored in the knowledge base of ES. It includes **forward chaining** (or goal driven control strategy); **backward chaining** (or data driven control strategy) and a hybrid strategy combining both these strategies. Forward chaining works from an

initial state of known facts to the goal state and backward chaining works from a hypothetical goal state to the facts perhaps in terms of subgoals. The subgoals are preconditions for the goal stated. If the hypothesis is not supported by facts, it tests for another goal state and so on in a predefined order of goals.

b. **The formation approach:** It involves forming a solution from eligible solution components stored in the knowledge base. It includes problem reduction (into subprograms), plan - generate - test (which generates all possible solutions, prunes inconsistent solutions and tests the remaining solutions), and agenda control. In agenda control a priority rating to each task in the agenda is assigned and the tasks are performed according to the assigned priority.

These techniques may be combined with other techniques for hierarchical planning, least commitment backtracking and constraint handling (Maher, 1986). Some other techniques available include inductive inference, metareasoning, ill structured problem and data handling etc.

2.3.3 Developing an ES

Developing an ES is a time consuming team work. Particularly, for developing a sophisticated ES an intensive and coherent effort is required. Knowledge engineers and domain experts work together to design an ES. The knowledge engineer develops the expert system, and the domain expert provides the

information for the knowledgebase. Hayes - Roth et al (1983) have identified five sequential stages in the development of ES, shown schematically in Fig. 2.4. Each stage is iterative in nature. Some of the stages are shown in Fig. 2.5.

2.4. ES Tools

A wide variety of development tools and environments are available for ES. These tools can be one of the general purpose programming languages or an ES shell. An ES shell is a set of ES development programs containing no knowledge about a problem, but can be 'taught' in a particular field or other. They contain all the modules required for ES. Filling of their hollow knolwdgebase makes them knowledge based expert system. Broadly, these ES tools are catagoriesed as:

1. Programming languages like PASCAL and C;
2. AI based exploratory programming languages like LISP and PROLOG;
3. ES shells like VIDHI which may be based on item "2";
4. High level ES programming environments like OPS5, ART, KES, NEXPERT, PC PLUS, RULE MASTER, CLIPS etc.; and
5. Mixed programming environments which allow the programmer to mix programming languages as in item "1" and item "2" with high level ES programming environment as in item "4".

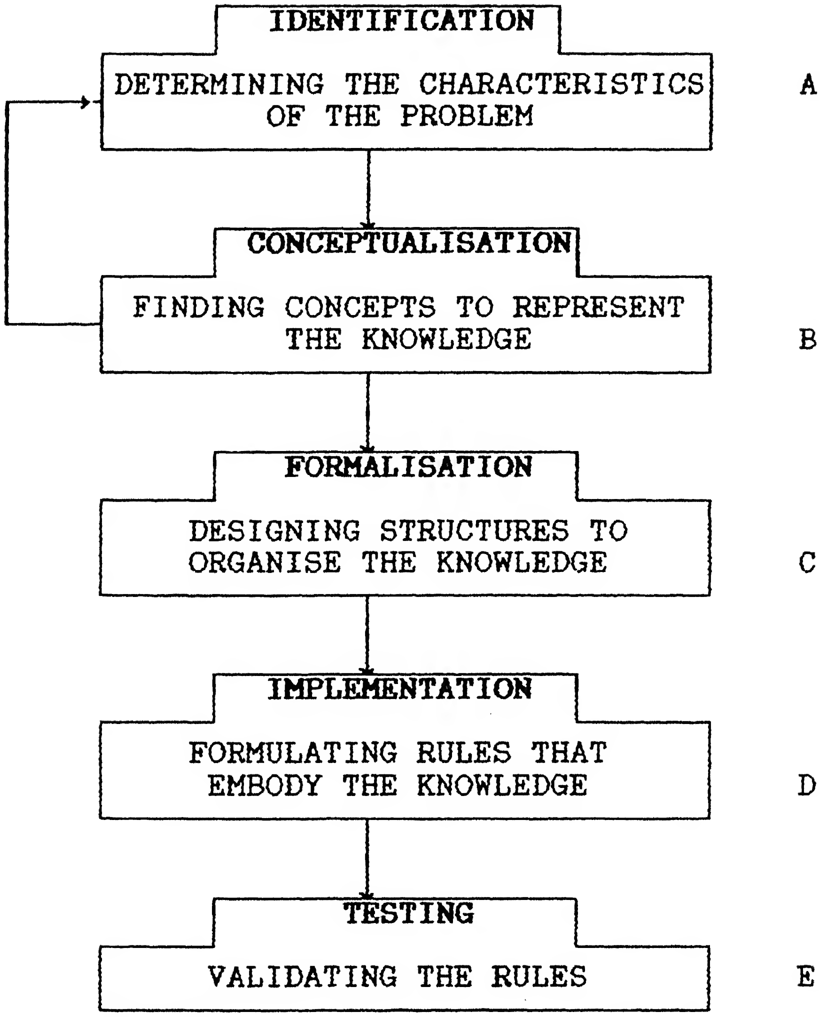

FIG. 2.4 : FIVE STAGES OF ES DEVELOPMENT
(Adapted from Hayes - Roth et al, 1983)

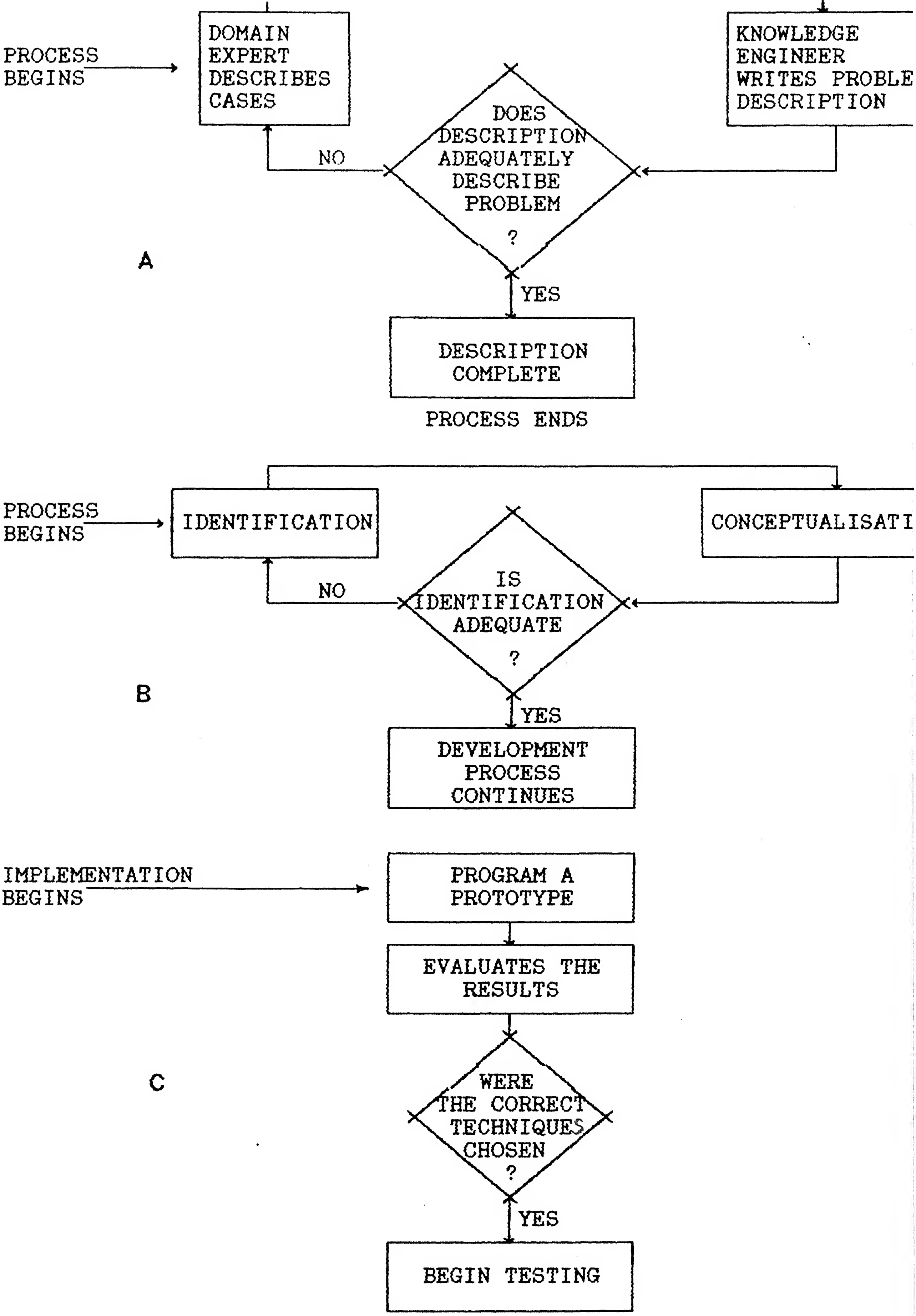

FIG. 2.5 : ITERATIVE STAGES OF ES DEVELOPMENT

2.5 Selection of ES Tool for Study

ES tools and shells available for the study incluc programming languages like FORTRAN, C, PASCAL etc., exploratoı programming languages LISP and PROLOG and an ES shell known as VIDHI a well as ES tools and environments including INSIGHT, PC PLUS RULE MASTER, and CLIPS. However, the choice of the tools depend upon the hardware and software available in IIT Kanpur and th topic of the study.

Several computer programs and subroutines developed eithe in IIT Kanpur or outside for frequency analysis of hydrologi data are available for the study. They include a compute program FAP for frequency analysis of seasonal hydrologic dat using different distributions and methods of paramete estimation; a computer program for fitting normal distribution b methods of least square; programs for transforming data with non normal distribution to those of normal or near norma distribution; a number of computer programs for fitting differen distributions to hydrologic data (Kite, 1977); and a number o computer programs developed by NIH Roorkee, HEC programs of U Army Corps of Engineers, Hydrologic Engineering center of US Arm Corps of Engineers etc. Furthermore, the hardware and softwar facilities available in IIT Kanpur computing environment includ HP 9000-800 system with UNIX operating system; microVAX I computer system with VMS-VAX operating system and IBM PC systems Programming languages FORTRAN, PASCAL and C are available in al the three systems. Because of extensive library of program

available in FORTRAN language for the study, it was proposed to use expert system in a mixed environment so that these FORTRAN programs can be accessed as necessary.

A brief comparison of the capabilities of the ES tools available for the study shows that CLIPS has a variety of features very much suitable for IIT Kanpur computing environment and interaction with FORTRAN programs:

1. It is written in C language for which compilers are available.
2. It is possible to call functions from CLIPS for numerical analysis, database, graphics etc.
3. It is possible to embed a CLIPS program within programs written in C, FORTRAN etc. and vice versa.
4. There is no limitation on length of embedded program other than the memory of the computer.
5. It can be ported to all the three computer systems available for the study.

Hence CLIPS is chosen as the ES tool in a mixed programming environment with FORTRAN for developing an ES for frequency analysis, specially in a VAX VMS environment.

CHAPTER III

CLIPS : AN ES SHELL

3.1 Introduction

CLIPS ('C' Language Integrated Production System) is a rule based forward chaining mixed programming expert system shell developed by National Aeronautics and Space Administration (NASA), Houston, Texas. The source programs of CLIPS (Version 4.3) are written in 'C' language. The primary representation methodology is a forward chaining rule language based on Rete algorithm (Forge, 1982) for multiple pattern matching and conflict resolution.

CLIPS can be used to develop a KBES in two ways, either as an

1. Interactive executable element; or
2. Embedded executable element.

In both ways the primary methodology remains the same with a difference in command syntax and in string conversion (if embedded with a language other than 'C'). CLIPS provides a programming environment in which a **knowledge base*** can be created and **facts** can be added, removed or generated. The **inference engine** attempts to arrive at a conclusion based on the defined set of domain specific knowledge base and current state of facts.

[* It may be noted that bold letters used in this Chapter other than the titles indicate reserved terms in CLIPS.]

3.2 Basic Features

3.2.1 Knowledge representation

a. **Facts**: A fact represents a piece of information and is placed in a current list of facts, the **fact-list**. It acts as a pattern for matching the conditions of a rule in order to **fire** that particular rule. Any amount of facts can be stored in the fact-list limited only by the memory of the computer. Facts can be added **(asserted)** or removed **(retracted)** in two ways, viz., prior to execution and by the action of a rule firing. A fact can be a **number**, a **word** or a **string**.

A **number** is any field which consists only of numbers (Ø-9), a decimal point (.), a sign (+ or -), and optionally an (e) for exponential notation with its corresponding sign. The number of significant digits and round of error may also occur depending upon machine implementation.

A **word** in CLIPS is any set of characters that starts with an alphabetic character and is followed by zero or more letters (A - Z), number (Ø - 9), under scores (_) or dashes (-). The word ends with a space. Space or other special characters may not be included within a word and furthermore CLIPS is case sensitive.

A **string** is a set of characters that starts with a double quotes (") and is followed by one or more letter (A - Z), numbers (Ø - 9), underscores (_), dashes (-), spaces. or special character (any printable character). A string ends with double

quotes. Double quotes or other special characters may be embedded within a field by pressing a backslash (\) in front of the character.

CLIPS facts are of ~~of~~ free form (no reserved words or order). The first fact in the fact list is called **initial-fact** and it is asserted by the system during a **reset**.

Defining initial facts: With the **deffacts** construct, facts can be added to the initial fact list. The asserted facts can be treated as any other facts. The initial fact-list, including any defined deffacts, is always reconstructed after a **reset**.

Syntax:

```
(deffacts <name> ["<comment>"]
        (<<fact 1>>)
        [    .
             .
             .
        (<<fact n>>)])
```

where <name> is a word used to identify the set of facts. Comment is optional.

b. Rule: The primary method of representing knowledge in CLIPS is a **rule.** A rule is a collection of actions to be taken if the conditions are met. The conditions are patterns which act as constraints and also provide a way to describe how to solve a problem.

Defining rule: Rules are defined using **defrule** construct. Each rule in CLIPS must have at least one condition and one action. There is no limit other than the actual memory of computer to the number of conditions or actions, a rule may have.

Syntax:

```
(defrule <name> ["<comment>"]
     [  .
        .                                         ;Left-Hand Side
        .                                         ;    (LHS)
      (<<nth pattern>>)]
      =>
      (<<first action>>)
      [  .
         .                                        ;Right-Hand Side
         .                                        ;    (RHS)
      (<<mth action>>)]])
```

where <name> is the name of the rule, a word and "<comment>" is optional.

LHS of the rule is made up of a series of one or more patterns which represent the condition elements for the rule. There is always an implicit (automatically implied) **and** surrounding all the patterns on the LHS. However **or**, **not**, explicit **and** (provided explicitly) or **test** (for mathematical or logical constraints), or a combination of them can also be incorporated. A rule priority can be assigned to a particular

rule by providing, in addition to other patterns, a **declare** pattern which declares its **salience** value ranging from a highest value of 10000 to a lowest value of -10000 . Single field pattern or multiple field patterns can be generated by proper syntax.

In RHS a list of the actions to be performed when LHS of the rule is satisfied is given. The arrow (=>) separates RHS from LHS. An action in CLIPS can be of many types. It can create new facts by **assert** action, can call an external function, can **retract** a fact, **assert** a string and **bind** a variable to a number or word (assign a value to the variable). Opening or closing a file, stopping CLIPS by **halt**, formatting by **format** and mathematical or logical predicate functions are some of the more general actions that can be incorporated. Two important features of the RHS actions **if-then else** structure and **while** structure are defined as follows:

<u>Syntax:</u>
I.

```
    (if (<predicate function> [<< args ....>>]
    then
       (<<action 1>>)
           .
       (<<action n>>)
   [else
       (<<action 1>>)
           .
       (<<action m>>)])
```

II.

```
(while <predicate function> [<<args...>>])
    (<<action 1>>)
        .
    (<<action n>>))
```

All predicate functions available in CLIPS are given in CLIPS Manual Part I (Culbert, 1987).

3.2.2 Inference engine

The inference engine of CLIPS is rule based and as indicated earlier (Sec. 2.3.1 and Sec. 3.1) works on a forward chaining inference mechanism. It is based on Rete algorithm (Forge, 1984) for multiple pattern matching and conflict resolution.

3.2.3 Basic cycle of execution

Unlike a conventional language the starting and stopping points are not explicitly defined by the programs in CLIPS. The inference engine applies the knowledge (rules) to data (facts). The basic cycle of execution is as follows:

1. The knowledge base is examined to see if the conditions of any rule has been met.
2. All rules whose conditions are currently met are **activated** and placed on the **agenda** (a stack). Rules having higher priority are kept on the top of the stack and activated before the new rule. Rules having lower or equal priority remain below the new rule.
3. The top rule on the agenda is selected and its RHS actions executed.

As a result of RHS actions new rules can be activated or/and deactivated.

This cycle is repeated until all rules that can fire have done so, or until a so called rule limit is reached (Appendix.A.).

3.3 Embedded Application of CLIPS

CLIPS has an added advantage of being integrated with, **external functions**, and/or 'C', FORTRAN or Ada language. This capability makes its application very flexible and more suited to engineering applications where numerical computations are of great importance, and where computer programs in any of the languages are already available.

3.3.1 External function

An external function defined by the user for his specific use in problem solving can be either in 'C' or in the language within which CLIPS is being embedded. An external function can be used on both the LHS and the RHS of rules. Data can be passed to and from them.

These functions are defined under **usrfuncs** either in file **main.c** (in interactive mode) or any other user file (in embedded application). Within **usrfuncs** a call should be made to the **define - function** routine for every function, the user wants CLIPS to know about.

Example

```
usrfuncs ()
    {
     define_function("fun", 's', fun, "fun);
     define_function("dummy", 'i', my_dummy, "my_dummy");
     /* Additional define function statements could go here.  */
    }
```

In define_function:

First argument is a CLIPS name, a string representation of the name that will be used inside CLIPS rule.

Second argument is a return type as 'i'=integer, 'f'=float, 'c'= character, 's'=pointer to a character string, 'w'=pointer to a character word, 'u'=pointer to an unknown data type.

Third argument is a pointer to the actual function (an external declaration of the function may be appropriate). The first argument need not be same as the actual function name, and

Fourth argument is a string representation of the actual function name. The name must be exactly the same as the function name, namely the third argument.

User defined functions are searched before system functions and if it matches with one of the defined functions already provided, the user function will be executed in its place.

a. **Passing variable from CLIPS to external function** CLIPS actually calls the function without any arguments, though they are listed directly following a function name inside CLIPS rules. Instead the parameters are stored internally by CLIPS and can be accessed by calling the functions:

```
int num_arg ();
char rstring  (arg);
float    rfloat (arg);
int    runknown (arg);
int    arg;
```

A call to **num_arg** will return an integer telling how many arguments the function was called with.

A call to **rstring** returns a character pointer, and **rfloat** returns a floating point number. The parameters have to be requested one at a time by specifying the parameter position number as the argument to **rstring** or **rfloat**. If the type of the argument is unknown, **runknown** can be called to determine the type.

b. **Passing data from external function to CLIPS** An external function can pass data into CLIPS in two ways. It can return a value or can assert a new fact directly into the CLIPS fact - list. If the external function is to be used as predicate, it must return a floating point number; otherwise it can be a character, integer, word or unknown. Return values can be used

as predicates, bound to variables, or captured via pattern expansion. The return value does not have to be captured, but must be defined in CLIPS, and all external functions **must** return a value.

The other method of passing data, asserting a new fact directly into the CLIPS fact list, is done by calling the C function **assert**.

<u>Example</u>

```
FACT  * assert (string):
char  * string ;
```

Here string is a single string made up of floating point numbers and words. The return value is a pointer to a variable of type FACT.

<u>Example</u>

<u>Function call</u> (from external function)	<u>Resulting Fact</u> (in CLIPS fact-list)
assert ("Ram is a boy")	**(Ram is a boy)**

CLIPS provides some more advanced interface functions in which passing known variable types, accessing multifield variables, and building facts by **scratch** (to assert lots of facts) are possible (Culbert, 1987).

3.3.2 Embedded application:

a. General CLIPS is designed to be embedded within other programs. The embedding program can be a 'C' language program or

a FORTRAN program or an Ada program. In each case a main program is provided by the user which calls CLIPS like any other subroutine. The basic changes which are to be made to access CLIPS from main program in a different language are:

1. Object version of main.c is removed from the link list of programs given in the CLIPS manual (Culbert, 1987);
2. Prior to loading rules, CLIPS is to be called by calling the function **init_clips**
3. If CLIPS calls any external function, user functions must be defined.
4. Many of the capabilities which are available in interactive interface are available by proper function calls (Culbert, 1987).
5. If the embedding main program is in 'C' it must include two include statements, namely

```
# include <stdio.h>
# include "<clips.h>"
```

6. If the embedding language is other than 'C' two modifications are to be made. They are:

i. A special function must be defined to run CLIPS from FORTRAN (for Ada it is not needed). This function is **run_clips** and it converts the rule firing parameter before calling the CLIPS **run** function (vide Appendix A).

ii. Since each language and each machine passes parameters differently, and since the source code of CLIPS is written in 'C' language, every string passing from languages other than 'C' should be converted to a 'C' string and vice versa.

In the present study CLIPS has been embedded within a FORTRAN main program and so the FORTRAN-CLIPS interaction is dealt within greater details in Sec. 3.3.2. Since the package has been implemented on microVAX II only, the machine dependent features are not discussed. The commands for linking and execution for VAX-VMS environment are given in Appendix B.

b. Fortran - CLIPS interaction

For complete language mixing four basic capabilities are needed

1. A program in another language may be used as the main program and CLIPS can be called as needed for reasoning.
2. Facts can be asserted into CLIPS from other languages
3. CLIPS may call other functions written in any language from the RHS of a rule and may pass parameters to the function.
4. In languages which can provide a meaningful return value, external functions may be called from the LHS of a rule (i.e. used as predicate function).

The main program written in FORTRAN initializes CLIPS, loads the rule files, resets the process, asserts the facts and runs the program by calling specialised CLIPS functions. The functions are; **int_clips, load_rules, reset_clips, assert, run_clips** in a sequence. In order to load a rule file or to assert a fact the Fortran strings are to be converted into C-strings by calling a function **storec**. Similarly for reverse action, i.e., to pass a parameter from CLIPS to Fortran, the C-string is to be converted into Fortran by calling a function **loadc**.

Facts can be asserted to CLIPS from Fortran either as constraints or as variables. In both cases it must be a string (see Examples 1 and 2)

<u>example 1</u> Asserting a constraint fact:

<u>Function call</u>	<u>Asserts in the CLIPS fact-list</u>
FACT1 = 'Ram is a boy'	
CALL STOREC (FACT1, C_FACT)	
CALL ASSERT (C_FACT)	**(Ram is a boy)**

<u>example 2</u> Asserting a variable fact:

<u>Function call</u>	<u>Asserts in the CLIPS fact-list</u>
100 FORMAT (E 10.4)	
X=4.5	
WRITE (INT_FILE, 1000)X	
FACT1= 'X is ' // INT_FILE(1:10)	
CALL STOREC (FACT1, C_FACT)	
CALL ASSERT (C_FACT)	**(X is 4.5)**

The argument in **storec** function FACT1 is declared as integer and C_FACT as a character.

The FORTRAN program which interacts with CLIPS is listed in Appendix C.

Furthermore, to call a function from CLIPS which is not defined for CLIPS, it is necessary to define an external function or simply a function depending upon the purpose, viz., whether it is being called from CLIPS, or simply some parameters are to be passed from CLIPS to the function. An external function can also be defined (Subsec. 3.3.1) so that it can be used in two ways, viz., as a C external function or as a FORTRAN external function.

In order to pass the parameters to an external function a C interface routine is to be defined. The C-interface routine and **usrfuncs** both are put in the file **runclips.C** along with defining the function **run_clips** (vide Appendix A).

3.4 Other Features

CLIPS can be implemented on a number of machines and a number of special functions can be defined internally or externally or both. Again I/O routing system provided in CLIPS is quit flexible and allows a wide variety of interfaces to be developed and easily attached to CLIPS. Simple window system, batch system, and dribble system are some of the I/O routers that can be incorporated in CLIPS (Culbert,1987).

CHAPTER IV

AN ES FOR FREQUENCY ANALYSIS OF CONTINUOUS HYDROLOGIC DATA

4.1 Frequency Analysis of Hydrologic Data

4.1.1 Introduction

Hydrology is a natural science which deals with many complex processes associated with the hydrologic cycle. Hydrologic data observed from historical natural hydrologic phenomena are the only sources of information upon which quantitative hydrologic investigations are generally based. The collection of hydrologic data have been continuously expanding and are in ever increasing amount. However, the majority of hydrologic processes in nature and their outcomes are affected by a number of causative factors. They may also be stochastic by nature and governed by laws of chance. The various hydrologic variables that describe these random processes are therefore also governed by laws of chance and are random by nature.

The hydrologic analysis dealing with these random variables may be carried out by applying either a deterministic approach or a stochastic approach or a combination of these two. Deterministic approaches in hydrology apply principles and laws of physical sciences such as fluid mechanics, thermodynamics etc. However, it is not possible to include all or even a sufficient number of causative factors in deterministic cause and effect relationships in hydrology. Therefore, deterministic approaches in hydrology, e.g., dynamic flow equation, suffer from major

disadvantages of being highly subjective, losing some of the information and having no associated probability level (Yevjevich, 1972; Kite, 1977). Moreover a function of a random variable is also a random variable and so to draw inferences from observed data statistical and probabilistic methods are to be applied.

Probability theory is the mathematical discipline that deals with the measure of chance or likelihood of a random variable while statistics deals with computation of statistical measures using data which are observations and/or measurements (Chow, 1964). Hydrologic data can suitably be expressed in statistical terms and be treated with probability theory as they are highly erratic and commonly stochastic in nature. Statistical analysis interprets the hydrologic observations and attempts to extract maximum information from them. Then the hydrologic information can be presented in condensed form as graphs, tables of numbers and mathematical techniques can be applied for decision making in ater resources planning, conservation, development and control.

An important problem in hydrology deals with interpreting a ast record of observations of hydrologic variables in terms of heir future probabilities of occurrence. These variables can be loods, droughts, storages, rainfalls, water qualities, waves etc. olution to this problem may be sorted out in two ways either aphically or statistically. The plotting of the observations of e variables/the number of occurrences with respect to ranges of lues generally produces some kind of pattern. This pattern may used to extend the available data to estimate the design value. a large number of observed data are available covering a period

of record much larger than the return period for design, the design value can be estimated directly from the sample. The graphical processes are simple, visually presentable and assume no prior distribution. However, generally it is not possible to obtain such a long length of data and furthermore the procedure is highly subjective, empirical, nonunique and therefore not compatible with other phases of engineering design.

In the statistical method, **frequency analysis** is used to fit a theoretical probability distribution function (pdf) to the sample data. Statistical parameters of the pdf are estimated and in turn they are used to estimate the design value. This method takes care of the stochastic behaviour of hydrologic processes upto some extent and is more compatible with various phases of engineering and design.

4.1.2 Problems in frequency analysis of hydrologic data

There is no general agreement among hydrologists on the specific choice of any particular theoretical distribution for frequency analysis of a given hydrologic variable at a given site. Frequency analysis, being a data dependent technique, is subject~~ed~~ to many constraints, limitations and assumptions also. Therefore the solution proposed often incorporates heuristics based on subjective judgment, experience, and common sense. The reliability of frequency analysis of hydrologic data depends upon many factors. Some of them are discussed as follows:

a. Data Frequency analysis of hydrologic data is based on certain assumptions about the input data. In routine analysis the data

are assumed to be consistent, homogeneous and independent. However in actual practice these assumptions are seldom satisfied.

In fact, the maximum floods are seldom, if ever, measured due to uncertainty in occurrence time and lack of proper gauging facilities at the proper time and place. As a result, they are estimated from the extrapolation of rating curves and other techniques. These estimates cause inconsistency in data with possible large errors. Due to urbanisation, industrialisation, changes in instrumentation and observation, and construction of new hydraulic structures for flood protection or irrigation, the data become nonhomogeneous. Nonhomoginity may also result from the fact that the observations taken for a particular hydrologic variable may be due to more than one distinct cause and effect relationships. Again, for example if two flood peaks resulting from the same meteorological disturbances are recorded, the assumption of independence in the hydrologic data becomes questionable.

Other errors resulting from data are due to random and systematic measurement errors and due to missing data. Missing data are often estimated by correlation and regression analysis of available data at other sites and/or other hydrologic variables. Proper knowledge of the transferability and similarity of the involved processes is required for such interpolations. Otherwise, the statistical analysis may be adversely affected, often severely.

b. Variable Type Mostly hydrologic variables are continuous but

the observations are, due to the least count of the instrument and sampling in time, discrete values. This causes loss of information. Some rare events occurring in the intervals between observations may be excluded from the analysis due to their non-availability, and this can seriously affect the reliability of the frequency analysis.

c. **Sample size** In most cases of hydrologic variables, the underlying processes are of geologic age during most of which no records have been collected. Also it is not possible to obtain a sample of very large length due to economic and other constraints. Statistical parameters may be calculated from a sample of finite, too small length. These estimates are then utilised to fit a theoretical distribution to observed data. The associated uncertainties can be evaluated (Stendinges, 1980) using knowledge derived from subjective judgment and from past experience. The appropriate choice of sample sizes for different events are site and events specific, and are often based on common sense.

d. **Variations:** Some hydrologic variables occur in clusters of very low or very high values. For example, a particular season may be having very small monthly precipitation while another ~~seasons~~ season may have large values. In this case average monthly precipitation of each year may give a misleading result. These clusters are to be identified and then described statistically. Otherwise the probability of monthly values being in an interval around an average monthly value, may differ from the probability of those particular monthly values belonging to well defined clusters of very low or very high value (Yevjevich, 1972).

e. Identification of outliers: An outlier in a set of data is defined as an observation or subset of observations which appears to be inconsistent with the remainder of that set of data. The inconsistency can be interpreted as the observation being either significantly higher (outlier) or lower (inlier) at the high end compared to the values indicated by the rest of the data. If the observations is significantly higher or lower than expected in the low end, it will be termed inlier or outlier, respectively. Presence of these outliers or inliers seriously distort the estimation of distribution parameters. For example, the estimation of skew parameters in probability distribution such as log Pearson type 3 is greatly affected by the presence of low outliers. Since the appearance of outliers may be due to several reasons, e.g., application of inappropriate distribution function to historical data, nonstationarity of the time sequences of major floods (as, long term trends), abnormal diverse causes of rainfall, inaccuracy of measurements etc., it is not very easy to correctly identify them.

Attempts have been made with little success for detection of outliers and inliers (Singh, 1987; WRC, 1981) but still it is a matter of subjective judgment of the analyst that whether he identifies some observations for special scrutiny in a conventional flood frequency analysis apparently having outliers.

f. Specific problem in fitting a pdf: Apart from the factors mentioned above there are many more constraints. e.g., nonstationarity with respect to the process involved etc.,which are very difficult to be taken care of by mathematical rules. Two basic questions need to be answered while fitting a particular

distribution to a given set of hydrologic data. These questions may be stated as:

1. Which of the many distributions available is the 'true' distribution for the population ? and
2. Which are the most appropriate methods of parameter estimation and goodness of fit test ?

The validity of the first question is due to four factors:

i. The data may follow a particular distribution naturally.
ii. Sample data available are usually for relatively low return periods (i.e.,around the center of probability distribution) while the one required to be estimated is generally of larger return period (i.e. in the tail region of the distribution).
iii. If the sample is derived from a population of mixed distributions one particular pdf will not fit to it.
iv. Many distributions have similar shape in their centre but differ widely in the tails. Thus, it is possible to fit several distributions on the sample data and end up with several different estimates of a T year event.

Second question arises due to fact that a number of statistical parameter estimation methods and goodness of fit tests are available, none of which has an absolute superiority or universal validity. For example, parametric and nonparametric

goodness of fit tests are based on different assumptions about the statistics of population (Kite, 1977) and thus the reliability of these tests also depends upon the validity of these assumptions.

These problems, if not dealt with proper attention and with an intelligent knowledge rich heuristic approach, will often lead to misleading results. Hence domain specific human expertise can be effectively utilised to address these problems.

4.1.3 Expert knowledge related to frequency analysis of continuous hydrologic variables

Frequency analysis of hydrologic data is a vast domain. Being in use from years it has gathered some amount of expertise in terms of heuristics. This section deals with some of the expertise, available for frequency analysis of continuous hydrologic variables in general and flood frequency analysis in particular. Some basic ideas which leads a hydrologist to reach a final conclusion about selection of a particular type of distribution, can be listed as:

a. Seasonality, stationarity and normality of data series: Seasonal variations must be considered in the selection of pdf. To take care of seasonality, different distributions which may be suitable for that season only should be fitted. Nonstationarity of data with respect to time and space can be detected by recognising the trends of data using time - series analysis and by observing changes in environmental conditions.

b. Outliers: Presence of outliers should be tested right in the beginning of the analysis. If the frequency plots show the

possibility of the presence of outliers, standard tests (Singh 1987: Chow et al, 1988) should be adopted. Test for inliers and outliers for a normal distribution (Table 4.1) should be conducted and if they are present then data should be transformed to normal distribution perhaps using power transformation (Singh, 1987) and method of least squares is applied for parameters estimation. If the series is an exceedence series then data are transformed to negative exponential series but method of parameters estimation remains the same. i.e., method of least squares.

c. **Mixed distribution**: There is always a chance of confusion about combination of two distinct distributions i.e. mixed distribution. A simple investigation about the causes of a particular hydrologic event and a frequency plot can provide a rational answer. If it is a mixed distribution the frequency analysis can follow the procedures discussed in Singh (1974) and Hawkins (1974).

d. **Concept of parsimony**: Yevjevich (1972) as well as Haan (1977) recommended the selection of pdf having not more than two or three parameters. From mathematical point of view the more the number of parameters, the better it will fit a set of data. However, if number of parameters is more, reliability of estimated parameter becomes less. Thus an optimisation between flexibility of distribution and reliability of parameters should be achieved. Other things being equal, smaller number parameters are to be estimated leading to their parsimony.

e. **Suitability of distribution for different data** Selection of pdf largely depends upon the type of hydrologic variable, viz., flood, drought, rainfall etc. For low flows or droughts, extreme

Table 4.1

Test values of outlier and inlier departures.

*	SL	Outlier Inlier	Low 1 15-100	Low 2 20-100	Low 3 25-100	Low 4 30-100	Low 5 40-100
1	<0.01	Inlier	<-0.689	<-0.495	<-0.412	<-0.363	<-0.327
	>0.99	Outlier	>1.029	>0.643	>0.498	>0.418	>0.368
2	<0.05	Inlier	<-0.532	<-0.369	<-0.303	<-0.264	<-0.237
	>0.95	Outlier	>0.681	>0.421	>0.337	>0.285	>0.253
3	<0.10	Inlier	<-0.441	<-0.299	<-0.243	<-0.211	<-0.188
	>0.90	Outlier	>0.503	>0.321	>0.254	>0.217	>0.193
4	<0.20	Inlier	<-0.318	<-0.209	<-0.167	<-0.143	<-0.127
	>0.80	Outlier	>0.297	>0.197	>0.159	>0.137	>0.123
5	<0.30	Inlier	<-0.221	<-0.141	<-0.110	<-0.093	<-0.082
	>0.70	Outlier	>0.161	>0.112	>0.092	>0.081	>0.073
6	<0.40	Inlier	<-0.132	<-0.080	<-0.060	<-0.050	<-0.043
	>0.60	Outlier	>0.052	>0.043	>0.037	>0.034	>0.032
			High 1 15-100	High 2 20-100	High 3 25-100	High 4 30-100	High 5 40-100
1	<0.01	Outlier	<-1.054	<-0.654	<-0.511	<-0.429	<-0.377
	>0.99	Inlier	>0.679	>0.488	>0.407	>0.358	>0.323
2	<0.05	Outlier	<-0.683	<-0.433	<-0.341	<-0.290	<-0.256
	>0.95	Inlier	>0.529	>0.369	>0.300	>0.263	>0.235
3	<0.10	Outlier	<-0.500	<-0.322	<-0.258	<-0.221	<-0.195
	>0.90	Inlier	>0.438	>0.299	>0.241	>0.209	>0.186
4	<0.20	Outlier	<-0.295	<-0.197	<-0.161	<-0.139	<-0.124
	>0.80	Inlier	>0.317	>0.209	>0.166	>0.143	>0.126
5	<0.30	Outlier	<-0.159	<-0.112	<-0.094	<-0.082	<-0.074
	>0.70	Inlier	>0.221	>0.140	>0.110	>0.093	>0.082
6	<0.40	Outlier	<-0.051	<-0.043	<-0.039	<-0.035	<-0.032
	>0.60	Inlier	>0.132	>0.079	>0.060	>0.050	<0.043

Notes: 15-100,..., and 40-100 denote the range of sample size n in years; * denotes window; SL - significance level

value type 3 or Pearson type 3 distributions (Kite, 1977) will be suitable. For floods lognormal, Pearson type 3 and extreme value type 1 and extreme value type 3 distributions are suitable. For rainfall, extreme value type 1 distribution is preferable (Chow, 1964). Similarly for exceedence series, negative exponential distribution is well suited.

f. **Parameter ranges for different distributions:** Statistical parameters like descriptors of central tendency, dispersion and peakedness are most commonly used for preliminary selection of pdfs.

Hydrologic variables are generally asymmetrically distributed. Therefore, selection of the proper set of statistical parameters are important for identifying the "true" pdf. Generally arithmetic mean is believed to be the best central value parameter. The coefficient of skewness often acts as regionalisation parameter and becomes very important when mean and standard deviation have small variation. The coefficient of peakedness or kurtosis is also useful for determining the best fit pdf. Often a combination of skew and kurtosis determines the pdf type. A guide to the selection of pdf in terms of kurtosis and skewness of the sample data is given in Fig 4.1 (Raudkivi, 1979). Some criteria for the initial selection of pdf based on estimated sample statistics are presented in Table No. 4.2.

The following notations are used:

μ = population mean

σ = population standard deviation

C_s = coefficient of skewness for population

$CV = \sigma/\mu$ = coefficient of variation

C_k = coefficient of kurtosis

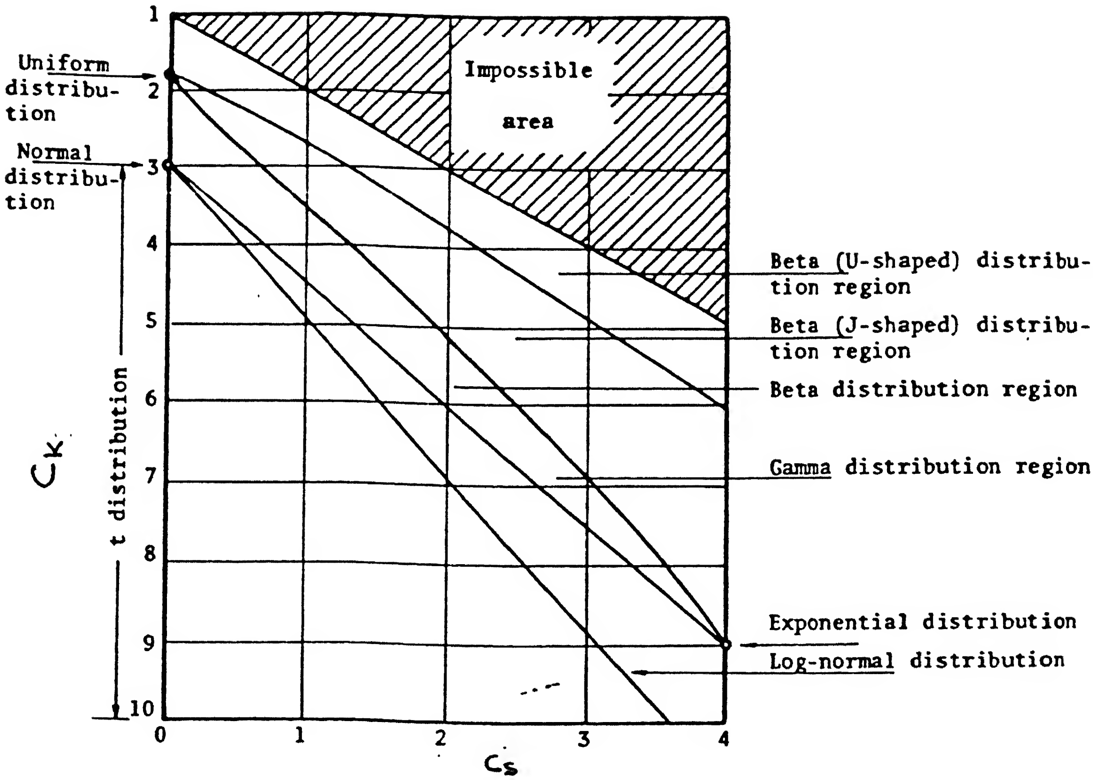

Fig. 4.1 Regions in (C_S, C_K) plane for various distributions, after E.S. Pearson. From Hahn and Shapiro (1967).

Table 4.2

Theortical and Heuristic Parameters Ranges

for Different Frequencies Distributions.

case nos.	Theortically	Heuristically	Suggestion
1.	$C_s = 0.0$	$-0.5 < C_s < 0.5$	
	$CV < 0.4$	same	
	$C_k > 1.0$	same	
	and		
	i. if $C_k < 2.0$	if $C_k < 2.25$	Uniform distribution
	ii. if $C_k = 3.0$	if $2.0 < C_k < 4.0$	Normal distribution
	iii. if $C_k > 3.0$	if $4.0 < C_k < 10.0$	t distribution
2.	$C_s < 0.0$	$C_s < -0.5$	Pearson type III or Lognormal type III can be fitted. Priority may be given to Pearson type III on the basis of parsimony of parameters. ?
3.	$C_s > 0.0$	$C_s > 0.5$	
	i. $C_k < (1.25C_s + 1.0)$ or $C_k < 1.0$	same	No distribution (impossible region)
	ii. $C_k = 2C_s + 3.0$	$1.25C_s + 3 < C_k < 2.25C_s + 3$	Lognormal II distribution

contd

iii. $C_s = 1.14$ and $C_k = 5.4$	$0.64 < C_s < 1.64$ $4.5 < C_k < 6.5$	Extreme value type I distribution
iv. $C_k = 3.0+6CV$ Where, $CV = 1/\sqrt{N}$	$2+6CV < C_k < 4+6CV$ Where, $\frac{1}{\sqrt{N}} -0.2<CV<\frac{1}{\sqrt{N}} +0.2$	Gamma distribution
v. $C_s = 4.0$ $C_k = 9.0$	$3.5 < C_s < 4.5$ $8 < C_k < 10$	Exponential distribution
vi. $(1.0625C_s+1.75)< C_k<(1.8125C_s+1.75)$	same	Beta (j) distribution
vii. $(C_s+1.0) < C_k<(1.0625C_s+1.75)$	same	Beta (u) distribution
viii. $(1.25C_s+3.0) < C_k<(1.8125C_s+1.75)$	same	Beta or Gamma distribution

Note : Alternatively

1. Input data can be log transformed to a Y series and then transformed data can be tested for Normal,Pearson type III or External type I disributions.

g. **Method of parameter estimation:** Theoretically, method of maximum likelihood estimate is the best method (Chow et al, 1988) for parameters estimation if the sample size is more than 25. However, if outliers are present method of least squares or special WRC procedures (Chow et al, 1988) must be used for the parameter estimation. Stedinger (1980) has discussed about specific methods of parameter estimation for lognormal distribution based on length of data and coefficient of Skewness. Similarly specific methods for parameter estimation for Pearson type 3 (Lall, 1987), and Log Pearson type 3 (Rao, 1987) have been formulated.

h. **Transformations:** Haan (1977) suggests that transformations are useful in the case of bounded distributions. Transformations are applied to normalise the data. Depending upon the coefficient of skewness a particular type of transformation is chosen. For a positively skewed data log transformation or n^{th} root transformations is needed. For negatively skewed data inverse Pearson or n^{th} root transformation is needed. Two step power transformation (Gupta, 1989) is the most powerful transformation.

Though this discussion does not provide an answer to all questions and uncertainties raised earlier, it clearly suggests a set of rule based procedures that can be utilised in the frequency analysis of continuous hydrologic variables. It clearly states some of the associated problems and suggests some solution procedures.

An extension of these rules, based on more vigorous evaluation and specific expertise of knowledgeable experts in this

field can be introduced as knowledge based for solving these problems. This can actually lead to a comprehensive expert system for frequency analysis of hydrologic data.

4.2 Structure of FACHVES Program

4.2.1 Introduction

Frequency Analysis of Continuous Hydrologic Variables Embedding Expert System (FACHVES) program is an interactive menu based FORTRAN program with an embedded expert system to provide decision making support for frequency analysis of continuous hydrologic variables. The program consists essentially of four parts, a main program, subroutines, ES interface and a rule based knowledge base.

4.2.2 Environment

FACHVES programs have been developed, implemented and tested in VAX - VMS environment. However it can be implemented on HP 9000-800 UNIX environment or on IBM PC with slight modifications.

4.2.3 Main program

The main program of FACHVES is a menu based program (Fig. 4.2) consisting of one master menu and eight submenus.

The master menu can call any of the eight submenus depending upon the option of the user. The submenus are basically for preliminary analysis, method of analysis, choice of distribution, transformations, goodness of fit test, plotting of results and

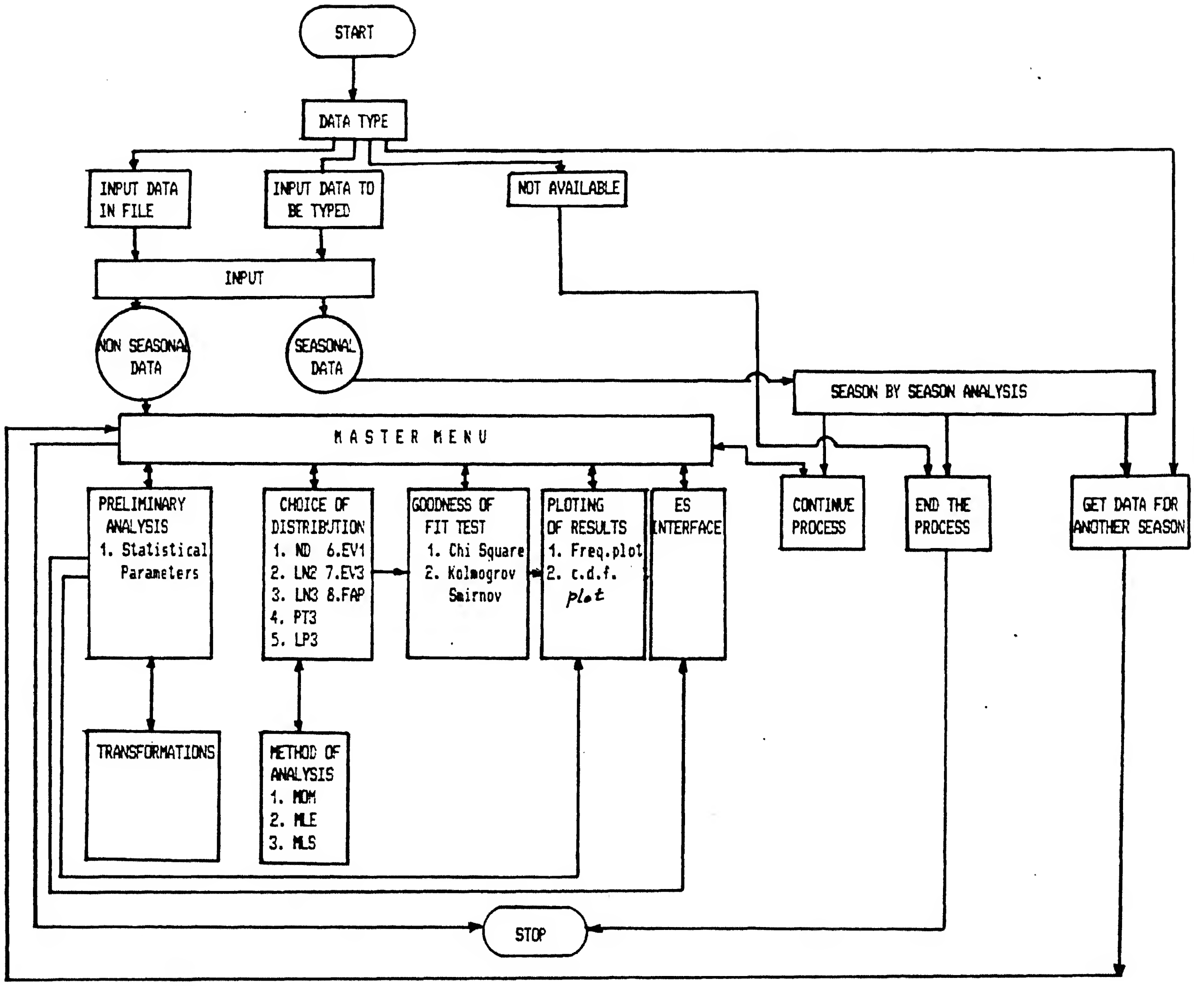

Fig. 4.2 Schematic Flow Chart of Main Program

ES interface. Each submenu consists of several options and depending upon user's choice, the desired subroutine can be called. However, the user is supposed to create an input file or give input data interactively before calling the master menu for the first time. Options for seasonal data analysis is displayed in the master menu only when the data is seasonal. A listing of the main program is given in Appendix D.

4.2.4 Subroutines

In FACHVES several subroutines already available in FORTRAN for statistical analysis and some subroutines, e.g., for two step power transformation have been incorporated. They can be called either from main program or from other subroutines, as per requirement. The subroutines are basically for

a. fitting different distributions;
b. applying different goodness of fit tests;
c. transforming the data using different transformations;
d. using different methods of analysis;
e. linking FORTRAN program to CLIPS, and
f. converting strings from FORTRAN to C and vice versa.

The subroutines are programmed in such a way that they perform one or a combination of the above functions. The limitations in using a subroutine, if any, are conveyed to the user interactively during execution. Table 4.3 lists the major subroutines used in frequency analysis.

One of the subroutines PTHELP acts as **ES interface** and it has

Table 4.3

List of Main Subroutines

Subroutine	Function
PARAM	Statistical parameter calculation
ND	Fits a normal distribution
LN2	Fits a lognormal 2 distribution and estimates T year events
LN3	Fits a lognormal 3 distribution and estimates T year events
PT3	Fits a Pearson type 3 distribution and estimates T year events
LP3	Fits a lognormal distribution and estimates T year events
TIE	Fits a Type 1 Extremal distribution and estimates T year events.
T3	Fits a Type 3 Extremal distribution and estimates T Year events.
FAP	Fits different distributions for seasonal data, calculates seasonal variations and tests the goodness of fit (Chi-square)
LOGTRAN	Trnsforms the data logarithmically
IPT	Transforms the data by Inverse Pearson transformation
SQTRAN	Transforms the data by square root transformation
TSPT	Transforms the data by two step power transformation
MLS	Fits a normal distribution using method of least square

contd

SER	Computation of the standard errors of events computed from various probability distributios compared to the observed event magnitude
CHISQRT	Chi-square test
KOLMG	Kolmogrov-smirnov test
PTHELP	ES interface ; calls ES program
STOREC	Converts a FORTRAN string in to C string
LOADC	Converts a C string in to FORTRAN string
ADVICE	Passes ES advice to main program

been incorporated as an optional calling subroutine from FORTRAN main program. The subroutine PTHELP links the main program with the CLIPS function calls and string conversion function STOREC as specified in Subsec. 3.3.2 (b). The suggestions given by ES is converted to FORTRAN strings, by LOADC function and stored in a subroutine ADVICE which subsequently passes them to the main program.

4.2.5 Knowledge base: The main function of the ES is to supplement the computational procedures specified in the main program with a number of advices for choosing the right alternatives for analysing the data. Without the ES knowledge base (KB) similar task could have been accomplished, however, only by going through an exhaustive process of all the available alternatives. The ES knowledge base helps to eliminate some of the non-feasible alternatives of the pdf fitting exercise. It also incorporates subjective judgments inbuilt in the KB, or provided by the user, to choose the right path of evaluation.

The ES is embedded in the main program and accessed by it during execution to accomplish the following tasks:

i. provide an interface for user supplied information in the decision making process.

ii. transfer appropriate parameters to the FORTRAN main program, based on the advice of the ES,

iii. advice on the adequacy of the data for statistical analysis

iv. advice on the initial choice of distribution functions based on estimated values of statistical parameters.

v. advice on the suitable choice of transformations based on

the estimated values of statistical parameters, and

vi. if the user feels he has outliers, then advice on the method to deal with outlier in the data.

The rule based ES's knowledge base is composed of a set of rules that reflects expert knowledge in this field. It also provides a vehicle for incorporating subjective judgment in the selection of procedures for statistical analysis, while guiding the user through various stages of decision making. As this study is of a preliminary nature with the goal of developing a framework for an ES for frequency analysis of continuous hydrologic variables, the rules developed are only exploratory in nature and are certainly not complete. The flow chart showing the various components and functions of the ES knowledge base is shown in Fig. 4.3.

Knowledge base is accessed during the execution of the main program. The type of inputs provided to the knowledge base directly from the main program includes the estimated values of the statistical parameters of the raw hydrologic data, and the length of the record. Based on these inputs, the ES guides the user through a series of steps, where existing patterns (rules) in the KB, are matched with the patterns constructed through user input and given input information. Matching or nonmatching of these patterns leads to either a set of parameters or control variables being passed on to the main program, or the testing of a subsequent rule.

Once all the existing rules compatible with the facts,

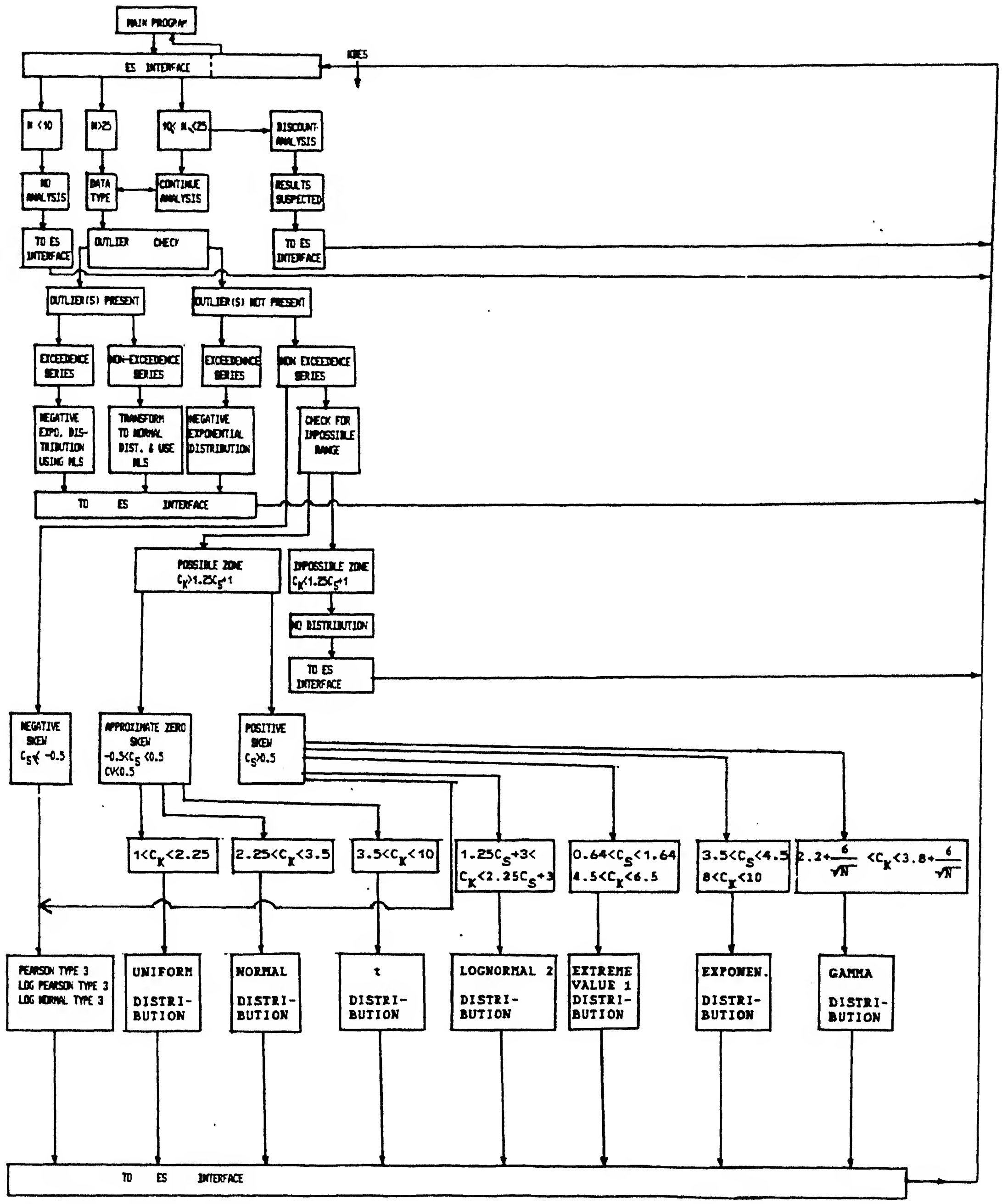

Fig. 4.3 Flow chart of Knowledge Base of ES.

asserted by the user or generated by the actions of rules have been tested, all the parameters and control variables as specified by the ES are transferred to the main program. Also a list of suggestions on the subsequent procedures of statistical analysis are displayed by the ES, to help the user. The ES knowledge base can be accessed by the user from the main program by using either the master menu or the submenu preliminary analysis, whenever user needs some advice.

4.3 Data Used

The data used to test the developed program are of two types; seasonal (monthly data along with monsoon,nonmonsoon and annual) and nonseasonal data. Two sets of seasonal data have been tested to verify the feasibility of this approach. These are:

A.Non seasonal data

a. Annual peak discharge of river Narmada (1951-1982)

b. Annual peak discharge of river Mahanadi at Hirakud(1946-1982)

B. Seasonal data

In seasonal data, monthly data along with monsoon, nonmonsoon and annual data have been considered as seasons for convenience of analysis. Therefore, 12 months and 3 seasonal sums in a year have been considered. The Hirakud flows for 1946 to 1982 have been used in the analysis.

The data used in the analysis are presented in Table 4.4.

Table 4.4

Input Data

```
THE ANNUAL PEAK FLOOD DISCHARGE AT MORTAKKA FOR NARMADA (1951-1982)
28,1,7,0
11127.0,13631.0,19521.0,33915.0,20746.0,11982.0,25023.0,31604.0,16135.0,
23438.0,18591.0,11338.0,19690.0,31604.0,27935.0,41691.0,18101.0,47851.0,
54003.0,36562.0,33278.0,17713.0,24354.0,29564.0,26232.0,22751.0,25662.0,
16602.0
```

```
   HIRAKUD FLOWS(ANNUAL) IN THM (1947-82)
 37,1,7,0
    6089.80   4771.50   4773.40   4271.20
    4344.30   2442.10   4667.90   3419.30   2037.10   3837.20
    4271.40   2413.20   4425.30   4645.00   4253.20   9377.60
    2058.90   3890.20   5500.50   1617.90   2275.90   4121.60
    2662.80   2638.10   4632.90   4913.10   2501.80   5449.00
    1821.20   4283.30   3226.40   3927.20   3689.40   1306.90
    4611.50   2708.20   2482.00
```

HIRAKUD FLOWS IN THM (1946-82)

555,15,11,0

289.5	1304.6	3054.5	880.7	327.7	100.4	48.0	25.9	25.9	19.0	9.9	3.7	5857.0	232.8	6089.8
8.5	1030.0	1745.7	1499.7	338.7	54.6	31.0	31.1	15.5	8.5	5.0	3.2	4622.6	148.9	4771.5
60.6	769.4	2005.4	773.8	943.7	104.8	55.4	24.1	17.5	9.3	4.9	4.5	4552.9	220.5	4773.4
62.5	652.6	1864.2	927.8	495.1	150.7	45.5	21.2	16.6	22.6	9.4	3.0	4002.2	269.0	4271.2
83.0	1172.4	2049.1	719.6	107.3	43.0	25.0	17.7	11.0	14.6	99.5	3.1	4131.4	212.9	4344.3
36.9	226.8	1413.8	405.8	243.0	62.8	22.3	10.3	7.5	5.7	2.1	5.1	2326.3	115.8	2442.1
70.2	805.8	1483.7	1874.0	257.6	79.7	34.1	36.3	18.5	4.0	2.3	1.7	4491.3	176.6	4667.9
22.2	991.0	1555.8	580.7	169.0	68.0	19.2	9.2	1.7	0.8	0.7	1.0	3318.7	100.6	3419.3
35.0	222.9	527.6	1030.1	170.7	40.0	3.2	3.6	1.6	0.8	0.8	0.8	1986.3	50.8	2037.1
99.9	755.9	911.0	1287.0	617.1	102.5	21.1	17.1	14.0	5.9	3.2	2.7	3670.9	166.3	3837.2
303.0	1414.7	1618.5	534.7	245.3	74.1	10.8	11.8	10.7	25.1	13.8	8.9	4116.2	155.2	4271.4
12.0	593.8	1236.9	387.6	41.3	28.7	21.3	16.4	27.1	44.9	1.6	1.6	2271.6	141.6	2413.2
44.4	1433.2	886.2	1240.9	599.0	111.4	43.1	28.1	17.2	6.3	10.3	5.2	4203.7	221.6	4425.3
30.7	625.5	1816.7	1708.9	273.0	72.3	40.3	28.8	10.0	19.6	9.3	9.9	4454.8	190.2	4645.0
62.8	797.9	2219.7	495.0	453.9	76.8	40.9	32.0	40.1	13.4	12.3	8.4	4029.3	223.9	4253.2
675.7	2762.1	1896.3	2990.9	719.4	150.6	86.6	41.9	19.9	16.7	8.9	8.6	9044.4	333.2	9377.6
59.7	406.1	733.5	555.0	131.2	61.9	34.3	16.1	15.2	19.9	2.5	3.5	1905.5	153.4	2058.9
96.3	611.9	1352.3	1370.2	265.6	93.7	40.9	21.7	12.1	17.2	3.4	4.9	3696.3	193.9	3890.2
165.9	1566.6	2168.0	863.0	510.1	113.7	19.9	29.2	11.8	26.0	18.8	7.5	5273.6	226.9	5500.5
45.2	389.2	385.6	558.8	113.3	43.1	23.9	21.2	16.6	6.8	2.5	11.2	1492.6	125.3	1617.9

CONTD.

266.9	555.0	1044.3	202.9	90.0	37.7	33.4	11.4	6.8	8.3	11.8	3.4	2959.1	116.8	2275.9
58.7	689.8	2058.6	930.8	172.2	57.1	52.7	33.9	32.5	19.8	5.7	9.8	3910.1	211.5	4121.6
61.3	552.0	1393.0	337.5	176.5	55.6	31.0	17.1	12.4	11.9	5.1	9.4	2520.3	142.5	2662.8
34.7	603.5	1264.9	448.4	131.7	59.7	33.4	21.8	12.5	19.5	4.5	3.5	2483.2	154.9	2638.1
176.4	1305.1	1726.4	1016.8	241.5	56.8	35.0	23.6	18.8	14.8	8.6	9.1	4466.2	166.7	4632.9
579.0	1297.8	1680.5	850.9	309.7	87.2	45.5	25.4	19.5	10.2	3.9	3.5	4717.9	195.2	4913.1
14.9	530.6	896.6	683.8	175.8	86.7	55.2	18.8	17.3	10.0	8.6	3.5	2301.7	200.1	2501.8
11.6	1043.9	1492.8	1588.3	938.8	252.8	56.9	31.3	19.3	6.9	2.9	3.5	5075.4	373.6	5449.0
26.8	355.7	1047.7	123.2	147.1	53.6	30.4	12.5	9.8	7.3	3.6	3.5	1700.5	120.7	1821.2
46.3	846.3	1836.6	825.0	482.9	135.7	47.3	26.3	18.5	9.9	5.0	3.5	4037.1	246.2	4283.3
14.6	662.2	1600.5	774.4	82.9	38.1	23.6	11.1	6.8	6.2	2.5	3.5	3134.6	91.8	3226.4
198.7	966.3	1419.1	924.8	234.5	75.0	42.5	18.7	21.5	18.9	3.7	3.5	3743.4	183.8	3927.2
109.4	683.7	1857.5	701.4	154.4	73.8	49.9	23.3	23.8	6.2	2.5	3.5	3506.4	183.0	3689.4
20.5	283.7	769.6	81.5	120.2	33.0	24.2	9.2	6.8	6.2	2.5	3.5	1275.5	85.4	1306.9
278.6	1199.8	1012.9	1816.6	163.9	61.7	31.1	21.2	10.0	6.2	6.0	3.5	4471.8	139.7	4611.5
35.5	523.9	1009.3	634.0	228.6	60.9	33.1	24.3	25.8	26.8	2.5	3.5	2531.3	176.9	2708.2
31.2	254.6	1421.6	468.9	157.1	65.1	32.1	15.4	20.6	8.6	2.6	4.3	2333.4	148.6	2482.0

4.4 Results and Discussions

The developed program can be used in two ways:

a. Analysis for annual data (if data are nonseasonal)

b. Season by season analysis (if data are seasonal)

A. Analysis for annual data

Analysis of two sets of data have been done and the interactive session with the program for one of them has been reported completely in Table 4.5.

a. Analysis for annual peak discharge

The statistical analysis for Narmada river shows that data are positively skewed (0.959) and coefficient of kurtosis is 4.0343. The KBES suggests a distribution which can be a Pearson type 3 or log Pearson type 3 or log normal type 3. The suggestion is also confirmed by flow chart of KBES (Fig 4.3) and Table 4.2. However, a Pearson type 3 distribution has been selected. Chi-square statistics of 2.998 for 3 degree of freedom (χ^2 95 %, 3 DOF=7.81) indicates that a distibution with parameters shown in Table 4.5 is a good fit at 95% confidence level. Table 4.5 shows the whole session and the fitted distribution.

Similar analysis for Mahanadi river at Hirakud indicates a positively skewed data (1.15) having coefficient of kurtosis 6.612 and the KBES gives similar suggestion. However a Pearson type 3 distribution has been selected. Chi-square statistic of 5.165 for 3 degree of freedom indicates that the distribution with parameters shown in Table 4.6 is a good fit at 95% confidence level. The fitted distribution is shown in Table 4.6.

b. Season by season analysis

Season by season analysis has been presented for three seasons and the interactive session with the program for one of

Table 4.5

Interactive Session I
(For annual data of peak discharges of river Narmada)

```
************************************************
*      MASTERS THESIS OF O.P.MISHRA           **
*      UNDER  THE  GUIDANCE    OF             **
*      Dr.S.Ramaseshan and Dr.B.Dutta         **
*           at I.I.T.Kanpur                   **
************************************************
*    WISH   YOU    HAPPY COMPUTING            **
*           W I T H                           **
*         F A C H V E S                       **
************************************************
* *--* FREQUENCY-ANALYSIS WITH EXPERT ADVICE *--*
* ---------------------------------------------*
*****************************************
*     I N P U T    M O D U L E          *
* --------------------------------------*
Do you want to give input on the screen?
If yes, Type 1 ,If no, Type 2
    If you want to quit Type 3
2
--The input file should be named as TEST.INP
--         It must have following :
--       *   a TITLE with format 80A1
--             AND, unformated
--       *   NX--Number of data
--       *   NS--Number of seasons
--          (for annual series NS =1)
--       *   NPT--Number of points
--           NPT=NCLASS+1 where NCLASS is number of
               classes in which data can be divided
               for chisquare test. The class frequency
               should be equal to 5
--            (It can be 7/11/13)
--       *   Observational hydrologic data RX(I)
--       (Not more than 600;
                seasons(in row)/years(in column)
Have you got such an input file ? If yes,type 1
       If you want to quit ,type 2
       If you want to go to previous menu type 3
1
*--------------------------------------------*
*      MASTER-MENU                           *
*--------------------------------------------*
*   1.PRELIMINARY ANALYSIS                   *
*   2.CHOICE OF DISTRIBUTIONS                *
*   3.TESTS FOR GOODNESS OF FIT              *
*   4.PLOTTING OF RESULTS                    *
*   5.QUIT                                   *
*   6.HELP FROM EXPERT SYSTEM INTERFACE      *
*--------------------------------------------*
**********************************************
Type the serial number of desired option
1
*******************************************
*   $$$ PRELIMINARY ANALYSIS $$$          *
*-----------------------------------------*
*   1.STATISTICAL PARAMETERS              *
*   2.FREQUENCY PLOT                      *
*   3.C.D.F.PLOT                          *
*   4.TRANSFORMATIONS                     *
*   5.HELP FROM EXPERT SYSTEM INTERFACE   *
*   6.TO MASTER MENU                      *
*-----------------------------------------*
Type the serial number of desired option
```

```
1
 MEAN=  25382.2852
 VAR=   0.11461E+09
 SKEW=   0.95990E+00
 STDEV=   0.10706E+05
 KURTOSIS=   0.40343E+01  AVG.DEVIATION= 0.82410E+04
THIS ANALYSIS IS FOR LOGTRANSFORMED DATA
 MEAN=      10.0593
 VAR=   0.17221E+00
 SKEW=   0.96092E-02
 STDEV=   0.41498E+00
 KURTOSIS=   0.28240E+01
XMIN= 0.93171E+01  XMAX= 0.10898E+02
MEAN,AND,STDEV FROM SCREEN

10.0593

0.41498
FOR ORIGINAL DATA****
CHECK FOR OUTLIERS BY WRC METHOD
XMIN= 0.11127E+05  XMAX= 0.54063E+05
No.of high outliers =   0.
No.of low outliers=   0.
******************************************
*   $$$ PRELIMINARY ANALYSIS $$$         *
*----------------------------------------*
*   1.STATISTICAL PARAMETERS             *
*   2.FREQUENCY PLOT                     *
*   3.C.D.F.PLOT                         *
*   4.TRANSFORMATIONS                    *
*   5.HELP FROM EXPERT SYSTEM INTERFACE  *
*   6.TO MASTER MENU                     *
*----------------------------------------*
Type the serial number of desired option
5
*********************************************
*  $$ EXPERT SYSTEM INTERFACE $$       **
*---------------------------------------*
*   1.FOR PRELIMINARY ADVICE            *
*   2.FOR SECONDARY   ADVICE            *
*   3.FOR EXPERT ANALYSIS OF RESULTS    *
*   7.TO MASTER MENU                    *
*---------------------------------------*
Which help do you want?
Type the serial number of desired option
```

```
 1
 f-0     (initial-fact)
 f-1     (start)
 f-0     (initial-fact)
 f-1     (start)
 f-2     (mean 25382)
 f-3     (skew 0.95990002)
 f-4     (kurt 4.03429985)
 f-5     (stdev 10706)
 f-6     (length of data 28)
        W E L C O M E

            TO

        F A C H V E S

This  expert   system   is  designed  for  single
              peaked  continuous hydrologic variables   only

IF YOU DO NOT UNDERSTANT THE QUESTION ; Type nk

Type of data ---- type the first letter (e/a/s)
              Whether it is exceedence OR annual ?a
Do you think that outliers are present?  no
THIS ANALYSIS IS FOR LOGTRANSFORMED DATA
 MEAN=     10.0593  .
 VAR=   0.17221E+00
 SKEW=   0.96092E-02
 STDEV=   0.41498E+00
 KURTOSIS=   0.28240E+01
EMON=tDe98lFeEm0typMMaXandal0888E+8Lion of
              logtransformed data ?0.41498
You can fit
              1.a pearson type 3 distribution
              2.a logpearson type 3 distribution
              3.a lognormal 3parameter distribution
Gamma distribution is well suited
1.You can fit either one OR all available distributions
PROCESS IS BEING STOPPED;
              RESET THE PROCESS IF YOU WANT TO CONTINUE
              WITH OTHER SET OF DATA
Do you want any help ?  no
Suggestions passed should be taken in sequencial order
f-0     (initial-fact)
f-1     (start)
f-2     (mean 25382)
f-3     (skew 0.95990002)
f-4     (kurt 4.03429985)
f-5     (stdev 10706)
f-6     (length of data 28)
f-7     (type of data a)
f-8     (process continue yes)
f-9     (data is large enough)
f-10    (outliers are no)
f-11    (positive skew)
f-12    (stdev of logtran data 0.41497999)
f-13    (logtran val is 2.49346423)
f-14    (suggestion 15)
f-15    (esad PT3,LP3,LN3)
f-16    (suggestion 10)
f-17    (esad GD)
f-18    (stop)
f-19    (help no)
Now you are out of clips
Value passed =     15.00000
Name passed is =
PT3,LP3,LN3
```

```
*------------------------------------------------*
*       MASTER-MENU                              *
*------------------------------------------------*
*   1.PRELIMINARY ANALYSIS                       *
*   2.CHOICE OF DISTRIBUTIONS                    *
*   3.TESTS FOR GOODNESS OF FIT                  *
*   4.PLOTTING OF RESULTS                        *
*   5.QUIT                                       *
*   6.HELP FROM EXPERT SYSTEM INTERFACE          *
*------------------------------------------------*
**************************************************
VALUE PASSED IS    15.00000
NAME PASSED IS
PT3,LP3,LN3
Type the serial number of desired option
2
***********************************************
*  $$ METHOD OF ANALYSIS $$           *
*-------------------------------------*
*   1.METHOD OF MOMENTS               *
*   2.METHOD OF MAXIMUM LIKELYHOOI    *
*   3.METHOD OF LEAST-SQUARES         *
*   4.TO MASTER MENU                  *
*-------------------------------------*
Which method do you like?
Type the serial number of desired option
1
*****************************************
**   $$ CHOICE OF DISTRIBUTION $$      **
*---------------------------------------*
*   1.NORMAL                      *
*   2.LOG-NORMAL (2-PARAMETER)    *
*   3.LOG-NORMAL (3-PARAMETER)    *
*   4.PEARSON TYPE 3              *
*   5.LOG-PEARSON TYPE-3          *
*   6.EXTREME-VALUE TYPE-1        *
*   7.EXTREME-VALUE TYPE-3        *
*   8.FAP--PROGRAMS               *
*   9.TO MASTER MENU              *
*  10.TRANSFORMATIONS             *
*  11.METHOD OF ANALYSIS          *
*---------------------------------------*
If you want to transform the data**TYPE-10
If not,then which distribution? type the serial number
4
***********************************************
*  $$ TEST OF GOODNESS OF FIT $$             **
*-----------------------------------------*
*    1.CHI-SQUARE TEST                    *
*    2.KOLMOGOROV-SMIRNOV TEST            *
*    3.TO PLOTTING RESULTS                *
*    4.COMPARISION OF STANDARD ERROR      *
*      OF ESTIMATE FOR DIFFERENT          *
*        DISTRIBUTIONS                    *
*    5.TO MASTER MENU                     *
*    6.HELP FROM EXPERT-SYSTEM INTERFACE  *
*-----------------------------------------*
Type the serial number of desired option
1
```

```
CHI-SQUARE STATISTIC
********************

  2.998593

NO. OF DEGREES OF FREEDOM
*************************

          3
**********************************************
**   $$ PLOTTING RESULTS                   *
*--------------------------------------*
*    1.FREQUENCY PLOT              *
*    2.C.D.F.PLOT                  *
*    3.TO MASTER MENU              *
*----------------------------------------*
Type the serial number of desired option
```

6

```
3
*-----------------------------------------------*
*       MASTER-MENU                             *
*-----------------------------------------------*
*    1.PRELIMINARY ANALYSIS                     *
*    2.CHOICE OF DISTRIBUTIONS                  *
*    3.TESTS FOR GOODNESS OF FIT                *
*    4.PLOTTING OF RESULTS                      *
*    5.QUIT                                     *
*    6.HELP FROM EXPERT SYSTEM INTERFACE        *
*-----------------------------------------------*
***********************************************
VALUE PASSED IS     15.00000
NAME PASSED IS
PT3,LP3,LN3
Type the serial number of desired option
5
FORTRAN STOP
$
                        PEARSON TYPE 3 DISTRIBUTION

                             METHOD OF MOMENTS


       ALPHA      0.66981E+04               M1          0.25382E+05
       BETA       0.25547E+01               M2          0.11461E+09
       GAMMA      0.82707E+04               SKEW        0.12513E+01

 T,YEARS    2           5           10          20          50          100

 X     0.23245E+05 0.33094E+05 0.39632E+05 0.45825E+05 0.53713E+05 0.59543E+05
  T
 S     0.13833E+04 0.24156E+04 0.37362E+04 0.57524E+04 0.90896E+04 0.11960E+05
  T
```

Table 4.6

Fitted Distributions to Different Sets of Data

. HIRAKUD FLOWS (ANNUAL)

PEARSON TYPE 3 DISTRIBUTION

METHOD OF MOMENTS

ALPHA	0.10822E+04	M1	0.37935E+04
BETA	0.19977E+01	M2	0.23466E+07
GAMMA	0.16283E+04	SKEW	0.14150E+01

T,YEARS	2	5	10	20	50	100
X_T	0.34519E+04	0.48603E+04	0.58229E+04	0.67486E+04	0.79428E+04	0.88342E+04
S_T	0.18859E+03	0.32250E+03	0.48730E+03	0.75913E+03	0.12242E+04	0.16304E+04

.HIRAKUD FLOWS (MONSOON)

THREE PARAMETER LOGNORMAL DISTRIBUTION

METHOD OF MOMENTS

MEAN OF X	0.36383E+04
VARIANCE OF X	0.21487E+07
SKEW OF X	0.11436E+01
A	-0.37778E+03

T,YEARS	2	5	10	20	50	100
X_T	0.33948E+04	0.47026E+04	0.55580E+04	0.63718E+04	0.74219E+04	0.82113E+04
S_T	0.27421E+03	0.35519E+03	0.47614E+03	0.70147E+03	0.11334E+04	0.15421E+04

Contd.

.HIRAKUD FLOWS (NONMONSOON)

THREE PARAMETER LOGNORMAL DISTRIBUTION

MAXIMUM LIKELIHOOD PROCEDURE

TRIAL	A	F(A)
1	0.34464E+02	-0.14010E+00
2	0.25506E+02	-0.61615E-01
3	0.13168E+02	-0.26492E-01
4	-0.29256E+01	-0.11117E-01
5	-0.22773E+02	-0.45460E-02
6	-0.45801E+02	-0.18072E-02
7	-0.70666E+02	-0.69714E-03
8	-0.94548E+02	-0.25356E-03
9	-0.11371E+03	-0.86248E-04
10	-0.12250E+03	-0.20146E-04
11	-0.12561E+03	-0.52452E-05
12	-0.12628E+03	-0.10133E-05
13	-0.12462E+03	0.24438E-05
14	-0.12584E+03	-0.19073E-05
15	-0.12560E+03	0.35763E-06
16	-0.12445E+03	0.17285E-05
17	-0.12574E+03	-0.20266E-05
18	-0.12495E+03	0.11921E-05
19	-0.12413E+03	0.12517E-05
20	-0.12586E+03	-0.27418E-05
21	-0.12538E+03	0.71526E-06
22	-0.12600E+03	-0.95367E-06
23	-0.12644E+03	-0.65565E-06
24	-0.12407E+03	0.34571E-05
25	-0.12553E+03	-0.23246E-05
26	-0.12553E+03	0.00000E+00

A	-0.12553E+03
MEAN OF LN(X-A)	0.56953E+01
VARIANCE OF LN(X-A)	0.43202E-01
SKEW OF LN(X-A)	0.30518E-03

FOR GOOD USE OF THIS DISTRIBUTIONSKEW OF LN(X-A) SHOULD BE CLOSE TO ZERO

T,YEARS	2	5	10	20	50	100
X_T	0.17195E+03	0.22881E+03	0.26274E+03	0.29320E+03	0.33034E+03	0.35692E+03
S_T	0.10948E+02	0.14112E+02	0.18349E+02	0.24095E+02	0.33449E+02	0.41559E+02

them has been shown in Table 4.7.

The statistical analysis for the data for monsoon season at Hirakud indicates positively skewed data with coefficient of kurtosis 6.891. The KBES suggests Pearson type 3, log Pearson type 3 and lognormal type 3. However Chi-square statistic for 7 degrees of freedom comes 37.019 for Pearson type 3 distribution and 38.07 for log Pearson type 3. For lognormal type 3 this value is 13.387. Therefore the distribution selected is lognormal type 3 which indicates a good fit at 95% confidence level ($\chi_{95\%,7DOF}$= 14.1).

Similar analysis for nonmonsoon season indicates lognormal type 3 distribution to be a good fit with Chi-square statistic 4.508 at 95% confidence level with 7 degrees of freedom.

However for the month of the January the data is found to have an approximate zero?(0.0529), coefficient of kurtosis as 2.68 and coefficient of variation 0.401. The KBES suggests a normal distribution for the season. The session output is shown in Table 4.7. Chi-square statistic of 1.73 for 7 degrees of freedom indicates that the fitted distribution is a good fit at 95% confidence level.

4.5 Conclusions

The test runs of the program FACHVES with the above sets of data shows that the KBES is very helpful arriving at correct decisions based on informations available either from FORTRAN program or user for fitting a probability distribution to a given set of data. It also establishes the feasibility of the approach to combine a statistical computational process with the heuristic capabilities of an expert system. However the program is to be improved significantly if it were to be powerful expert system.

Table 4.7

Interactive Session II

(For seasonal data for the flows in January in river Mahanadi)

```
*************************************************
*       MASTERS THESIS OF O.P.MISHRA            **
*       UNDER  THE  GUIDANCE    OF              **
*       Dr.S.Ramaseshan and Dr.B.Dutta          **
*            at I.I.T.Kanpur                    **
*************************************************
*     WISH    YOU    HAPPY COMPUTING            **
*              W I T H                          **
*           F A C H V E S                       **
*************************************************
* *--* FREQUENCY-ANALYSIS WITH EXPERT ADVICE *--*
* ---------------------------------------------*
******************************************
*      I N P U T    M O D U L E          *
* --------------------------------------*
Do you want to give input on the screen?
If yes, Type 1 ,If no, Type 2
    If you want to quit Type 3
2
--The input file should be named as TEST.INP
--          It must have following :
--        *   a TITLE with format 80A1
--              AND, unformated
--        *   NX--Number of data
--        *   NS--Number of seasons
--          (for annual series NS =1)
--        *   NPI--Number of points
--            NPI=NCLASS+1 where NCLASS is number of
--              classes in which data can be divided
--              for chisquare test. The class frequency
--              should be equal to 5
--             (It can be 7/11/13)
--        *   Observational hydrologic data RX(I)
--        (Not more than 600;
--                seasons(in row)/years(in column)
Have you got such an input file ? If yes,type 1
       If you want to quit ,type 2
       If you want to go to previous menu type 3
1
For which season do you want to do the analysis ?
Type the serial no of the season

8
   THIS COMPUTATION IS FOR SEASON NO.          8
*----------------------------------------------*
*       MASTER-MENU                            *
*----------------------------------------------*
*    1.PRELIMINARY ANALYSIS                    *
*    2.CHOICE OF DISTRIBUTIONS                 *
*    3.TESTS FOR GOODNESS OF FIT               *
*    4.PLOTTING OF RESULTS                     *
*    5.QUIT                                    *
*    6.HELP FROM EXPERT SYSTEM INTERFACE       *
*    7.ANALYSIS FOR OTHER SEASONS              *
*----------------------------------------------*
************************************************
Type the serial number of desired option
```

```
1
*****************************************
*   $$$ PRELIMINARY ANALYSIS $$$        *
*---------------------------------------*
*   1.STATISTICAL PARAMETERS            *
*   2.FREQUENCY PLOT                    *
*   3.C.D.F.PLOT                        *
*   4.TRANSFORMATIONS                   *
*   5.HELP FROM EXPERT SYSTEM INTERFACE *
*   6.TO MASTER MENU                    *
*---------------------------------------*
Type the serial number of desired option
1
 MEAN=     22.0135
 VAR=   0.78063E+02
 SKEW=   0.52969E-01
 STDEV=   0.88354E+01
 KURTOSIS=   0.26858E+01  AVG.DEVIATION= 0.71443E+01
THIS ANALYSIS IS FOR LOGTRANSFORMED DATA
 MEAN=      4.6867
 VAR=   0.53217E+01
 SKEW=   0.43273E-01
 STDEV=   0.23069E+01
 KURTOSIS=   0.19095E+01
XMIN=-0.35667E+00  XMAX= 0.91461E+01

MEAN,AND,STDEV FROM SCREEN
4.6867

2.3069
FOR ORIGINAL DATA*****
CHECK FOR OUTLIERS BY WRC METHOD
XMIN= 0.36000E+01  XMAX= 0.41900E+02
No.of high outliers =   0.
No.of low outliers=   0.
*****************************************
*   $$$ PRELIMINARY ANALYSIS $$$        *
*---------------------------------------*
*   1.STATISTICAL PARAMETERS            *
*   2.FREQUENCY PLOT                    *
*   3.C.D.F.PLOT                        *
*   4.TRANSFORMATIONS                   *
*   5.HELP FROM EXPERT SYSTEM INTERFACE *
*   6.TO MASTER MENU                    *
*---------------------------------------*
Type the serial number of desired option
5
********************************************
*  $$ EXPERT SYSTEM INTERFACE $$       **
*-------------------------------------*
*   1.FOR PRELIMINARY ADVICE           *
*   2.FOR SECONDARY   ADVICE           *
*   3.FOR EXPERT ANALYSIS OF RESULTS   *
*   7.TO MASTER MENU                   *
*-------------------------------------*
Which help do you want?
Type the serial number of desired option
```

```
1
f-0     (initial-fact)
f-1     (start)
f-0     (initial-fact)
f-1     (start)
f-2     (mean 22.01399994)
f-3     (skew 0.052969)
f-4     (kurt 2.68580008)
f-5     (stdev 8.83539963)
f-6     (length of data 37)
       W E L C O M E

           TO

        F A C H V E S

This  expert   system   is  designed  for  single
               peaked  continuous hydrologic variables   only

IF YOU DO NOT UNDERSTANT THE QUESTION ; Type nk

Type of data ---- type the first letter (e/a/s)
               Whether it is exceedence OR annual ?s
Do you think that outliers are present?  no
NORMAL DISTRIBUTION IS WELL SUITED
1.You can fit one distribution for all seasons;OR
               2.You can fit different distributions for different
                 seasons;OR
               3.You can fit all available distributions for all seasons
PROCESS IS BEING STOPPED;
               RESET THE PROCESS IF YOU WANT TO CONTINUE
               WITH OTHER SET OF DATA
Do you want any help ?  no
Suggestions passed should be taken in sequencial order
f-0     (initial-fact)
f-1     (start)
f-2     (mean 22.01399994)
f-3     (skew 0.052969)
f-4     (kurt 2.68580008)
f-5     (stdev 8.83539963)
f-6     (length of data 37)
f-7     (type of data s)
f-8     (process continue yes)
f-9     (data is large enough)
f-10    (outliers are no)
f-11    (suggestion 6)
f-12    (esad ND)
f-13    (stop)
f-14    (help no)
Now you are out of clips
Value passed =     6.000000
Name passed is =
ND
*---------------------------------------------------*
*      MASTER-MENU                                  *
*---------------------------------------------------*
*   1.PRELIMINARY ANALYSIS                          *
*   2.CHOICE OF DISTRIBUTIONS                       *
*   3.TESTS FOR GOODNESS OF FIT                     *
*   4.PLOTTING OF RESULTS                           *
*   5.QUIT                                          *
*   6.HELP FROM EXPERT SYSTEM INTERFACE             *
*   7.ANALYSIS FOR OTHER SEASONS                    *
*---------------------------------------------------*
*****************************************************
VALUE PASSED IS    6.000000
NAME PASSED IS
ND
Type the serial number of desired option
```

```
2
***********************************************
*  $$ METHOD OF ANALYSIS $$           *
*------------------------------------*
*   1.METHOD OF MOMENTS               *
*   2.METHOD OF MAXIMUM LIKELYHOOD    *
*   3.METHOD OF LEAST-SQUARES         *
*   4.TO MASTER MENU                  *
*------------------------------------*
Which method do you like?
Type the serial number of desired option
1
*****************************************
**   $$ CHOICE OF DISTRIBUTION $$      **
*---------------------------------------*
*   1.NORMAL                       *
*   2.LOG-NORMAL (2-PARAMETER)     *
*   3.LOG-NORMAL (3-PARAMETER)     *
*   4.PEARSON TYPE 3               *
*   5.LOG-PEARSON TYPE-3           *
*   6.EXTREME-VALUE TYPE-1         *
*   7.EXTREME-VALUE TYPE-3         *
*   8.FAP--PROGRAMS                *
*   9.TO MASTER MENU               *
*  10.TRANSFORMATIONS              *
*  11.METHOD OF ANALYSIS           *
*----------------------------------------*
If you want to transform the data**TYPE-10
If not,then which distribution? type the serial number
8
It fits normal distribution with or withouttransformations
KOUNT=1---for normal distribution
KOUNT=2---for inverse pearson transformation
KOUNT=3---for squareroot transformation
KOUNT=4---for log-normal distribution(MLE)
KOUNT=5---for pearson distribution
KOUNT=6---for log-transformation(MOM)
KOUNT=7---to quit from fap
What is yours?
KOUNT=
1
 WISH YOU GOOD LUCK
************************************************
*  $$ TEST OF GOODNESS OF FIT $$             **
*-----------------------------------------*
*    1.CHI-SQUARE TEST                    *
*    2.KOLMOGOROV-SMIRNOV TEST            *
*    3.TO PLOTTING RESULTS                *
*    4.COMPARISION OF STANDARD ERROR      *
*      OF ESTIMATE FOR DIFFERENT          *
*        DISTRIBUTIONS                    *
*    5.TO MASTER MENU                     *
*    6.HELP FROM EXPERT-SYSTEM INTERFACE  *
*-----------------------------------------*
Type the serial number of desired option
```

```
CHI-SQUARE STATISTIC
********************

  1.72[illegible]

NO. OF DEGREES OF FREEDOM
**************************

         7
************************************************
**  $$ PLOTTING RESULTS                    *
*--------------------------------------*
*   1.FREQUENCY PLOT              *
*   2.C.D.F.PLOT                  *
*   3.TO MASTER MENU              *
*---------------------------------------*
Type the serial number of desired option
3
*-----------------------------------------------*
*      MASTER-MENU                              *
*-----------------------------------------------*
*   1.PRELIMINARY ANALYSIS                      *
*   2.CHOICE OF DISTRIBUTIONS                   *
*   3.TESTS FOR GOODNESS OF FIT                 *
*   4.PLOTTING OF RESULTS                       *
*   5.QUIT
*   6.HELP FROM EXPERT SYSTEM INTERFAC
*-----------------------------------------------*
*************************************************
VALUE PASSED IS    6.000000
NAME PASSED IS
ND
Type the serial number of desired option
5
FORTRAN STOP
$


ANALYSIS FOR NORMAL DISTRIBUTION
*********************************


SEASONAL AVERAGES
******************

     22.014

SEASONAL STANDARD DEVIATION
****************************

      8.835

SEASONAL COEFFICIENT OF SKEWNESS
*********************************

      0.053

CHI-SQUARE STATISTIC
*********************

      1.723

NO. OF DEGREES OF FREEDOM
**************************
```

CHAPTER 5

SUMMARY,CONCLUSIONS,SUGGESTIONS

5.1 Summary

Expert Systems are becoming very popular in several knowledge-rich domains and therefore also in water resources engineering. They are applied either to solve a problem making use of computerised human expertise; or to assist a problem solving procedure by proper advice based on heuristic expert knowledge, especially to help non experts solve a domain specific problem.

Out of several expert system tools CLIPS has been chosen the present study for its ability to communicate with exter programs written in other high level languages, enhan capability of handling numerical values, easy installati economy in terms of memory requirements and compatibility w several machines available for the study.

Frequency analysis is an important step in engineer design and planning, particularly in water resources engineeri where estimation of future hydrologic phenomena is based on p observations. Being a data dependent technique the reliability frequency analysis is always questionable. To fit a theoreti probability distribution function to a set of observations and draw conclusions therefrom, requires a combination of rout numerical/statistical computations and human expertise in the f of heuristics.

In order to address these problems a program, Frequency Analysis of Continuous Hydrologic Variables with an embedded Expert System (FACHVES) has been developed. This program was developed in FORTRAN for microVAX-VMS environment. It consists of methods for computerised statistical analysis supported by an expert decision support system. The framework consists of four parts: main program,subroutines, ES interface, and a knowledge base.

The developmental work completed so far is preliminary in nature and it provides the framework for further development of a more comprehensive computer program package. Two data sets from field were tested for evaluating the performance of the program developed so far.

Due to easy portability of CLIPS,C, and FORTRAN this package can be easily transported to other systems such as UNIX HP-9000 or IBM PC.

5.2 Conclusions

Frequency analysis of hydrologic data is a knowledge domain where lot of decisions are to be taken based on human intuition,experience,and subjective judgments.

The developed package for FACHVES was tested with two different data sets representing recorded streamflow data at two different sites in India. The results of this testing show the

proper functioning of the package at this stage of development. These results also established the feasibility of this approach for statistical analysis of the hydrologic data by a combination of computational procedures and expert advice.

Suggestions For Future Study

A huge amount of expert knowledge is required and available for frequency analysis of hydrologic data. This essentially preliminary study incorporates only a small fraction of these methods and expertise. Substantial additional study is necessary to make the developed package suitable for practical use. In particular, the following tasks are necessary:

1. Further enhancement of the knowledge base of the ES
2. Addition of other computational methods for statistical analysis
3. Enhancement of the graphical capabilities; and
4. Implementation of the package in other computing environments.

R E F E R E N C E S

1. Adeli, H., 1985. "Knowledge Based Expert Systems in Structural Engineering", Second Intl. Conference on Civil and Structural Engineering Computing, London, December

2. Ahmed, K. et al, 1989, Analysis of Expert System for Fluid Flow in Networks Using a Qualitative Reasoning, Proc. HYDROCOMP' 89, Elscier pp. 394-403

3. Arnold, U., B. Datta, and P. Haensehcid, 1989. "Intelligent Geographic Information Systems and Surface Water Modeling", New Trends in Surface Water Hydrology, IAHS Third Scientific Assenbly, USA

4. Arnold, C.L., 1986. Selection of Appropriate Water/Sanitation Treatment Technology in Developing Countries: An Expert System Application in 27, Not printed

5. Box, G.E.P. and Cox, D.R., 1964." An Analysis of Transformation", Journal of the Royal Statistical Society, Vol.B26, pp. 211-252

6. Chabris, C. F., 1988. "Artificial Intelligence and Turbo Pascal", Multiscience Press, Inc., Galgotia Publication, Delhi.

7. Chander, S., S. K. Spolia, and Arun Kumar, 1978. "Flood Frequency Analysis by Power Transformation", Journal of Hyd. Div., ASCE, Vol.104, No-Hy11, November

8. Chow, V.T., 1964. **Handbook of Applied Hydrology**, Mc-Graw Hill Book Co., New York, N.Y.

9. Chow, V. T., D. R. Maidment, L. W. Mayes, 1988. **Appplied Hydrology**, Mc-Graw Hill Publication, Singapore

10. Culbert, C.J., 1987. "CLIPS Reference Mannual" Version 4.0, NASA, Mission Support Directorate, Houston, Texas

11. Datta, B., and R.C.Peralta, "Expert System and Pattern Recognition to Identify Pollution Sources", World Water Issues in Evolution, ASCE, Vol.1, 1986

12. Datta, B., J.E.Beegle, M.L. Kavvas, and G.T.Orlab, 1990. "Development of an Expert System Embedding Pattern Recognition Techniques for Pollution Source Identification, Project Completion Report to United States Geological Survey, University of California, Davis

13. Delleur, J.W., 1988. "A Front end Expert System for the Calibrtion of SWMM runoff block", Annual Water Resoruces Conference, WRP & M Div., ASCE, Norfolk. va.

14.Duda, R. and Gashing, J.G., 1981. "Knowledge Based Expert Systems Come of Age", Byte, Vol.6, #9, September, pp. 238-279

15.Engman, E.T., et al, 1986. "An Expert System for Snowmelt Runoff Modelling and Forecasting, Proc. Water Forum, ASCE, pp. 174-180

16.Fayegh, D. and S.O. Russel, 1986. "An Expert System for Flood Estimation", in 27, pp. 174-181

17.Forgy, C.L.,1984. "Rete : A Fast Algorithm for the Many Pattern /.Many Object Pattern Match Problem" Artficial Intelligence, 19:17-37

18.Gasching, J. et al, 1981. "Development of a Knowledge Based System for Water Resources Problems", SRI Project 1619, SRI International, August

19.Giarratano, J.C., C.Culbert, G.Riley, and R.Savely, 1987."A Solution to the Expert System Delivery Problem", NASA/Johnson Space Centre, Artificial Intelligence Section FM7

20.Gupta, D.K. et al,1989. "Flood Frequency Analysis by Two-step Power Transformation", NIH Vol 70, November

21.Haan, C.T., 1977. **Statistical Methods in Hydrology**, The Iowa State University Press, Amer

22.Hahn, G.J. and S.S.Shapiro, 1967. **Statistical Models in Engineering**, Willey

23.Hayer-Roth, F.D. et al, 1983.,"**Building Expert Systems**", Addison-Wesley Publishers

24.Jackson,P., 1986., **Introducion to Expert System**, Addison-Wesley Publishers Ltd.

25.Johnston, D.M., 1985., "Diagnostic of Waste Water Treatment Process", Computer Application in Water Resources, ASCE, pp. 601-610

26.Kite,G.W., 1977., **Frequency and Risk Analysis in Hydrology**, Water Resources Publications, Fort collins, Colrado

27.Kostem, C.N. and M.L. Maher, 1986., **Expert Systems in Civil Engineering**, Proc. of Symposium Seattle, WA, U.S.A.

28.Lall, U., 1987., "Estimation of a Prior Distribution for the Bayesian Estimation of Pearson III Skews", in **Hydrologig Frequency Modelling**, ed. V.P.Singh, pp. 131-147

29.Levine, R.I., D.E.Darang, and B.Edelson, 1988., **A Comprehensive Guide to AI and Expert System : Turbo Pascal Edition**, Mc Graw-Hill Book Company, Singapore

30.Lindberg, S. and J.B.Nielson,1986., "modelling of Urban Storm Sewar System" in 48, pp. 687-696

31.Maher, M.L., 1986., **Problem Solving Using Expert System Techniques**

32.Markovic, D., 1965., "Probability Functions of Best Fit to Distributions of Annual Precipitation and Runoff", Hydrologic Pap 8. code. State Univrersity, Fort Collins

33.McDermott, D. and E.Charniak, 1985., **Introduction to Artificial Intelligence**, Addison-Wesley Publishers

34.Mishkoff, H.C., **Understanding Artificial Intelligence**, B.P.B. Publications, Delhi

35.Oswald, O., 1989."An Expert System for the Diagnosis of Tank Irrigation System", Ph.D. Thesis in Progress College of Engineering , Guindy, Madras, India

36.Perumal, M., P.Nirupama, and S.M.Seth, "Some Studies on Parameter Estimation Methods of Gumbel EV-I Distribution Using Monte Carlo Tests"

37.Ramaseshan, S., 1990., "Expert Systems for Issues in Water Resources ngineering", Keynote address ?

38.Rao, D.V., 1987., "Fitting Log Pearson type III Distribution by Maximum Likelihood", in Hydrologic Frequency Modelling, ed. V.P.Singh, pp. 395-406

39.Raudkivi, A.J., 1979., **Hrdrology**, Pergamon Press Ltd., Headington Hill Hall, Oxford, England

40.Rich, E., 1983., **Artificial Intelligence**, Mc Graw-Hill Book Co.,Singapore

41.Savic, D.A. and S.P.Simonoic, 1989., "Introduction of Fuzzy Reasoning in Intelligent Decison Support for Reservoir Analysis, in 2, pp. 404-414

42.Simonovic, S.P., 1989., "Expert System Decision Support for Surface Water Quality Data Management, in 2, pp. 383-393

43.Singh, K.P., and R.A.Sinelair, 1974., "Two Distribution Method for Flood Frequency Analysis, Journal of Hydrology Div., Proc. ASCE, Hy. 1, Vol. 98, pp. 29-44

44.Singh, K.P., 1987., "Flood Data, Underlying Distribution, Analysis, and Refinement", Hydrologic Frequency Modelling, ed. V.P.Singh, pp. 227-241

45.Singh, K.P., 1987., "Development of a Versatile Flood

Frequency Methodology and its Application to Flood Series from Differnt Countries", Hydrologic Frequency Modelling, ed. V.P.Singh, pp. 183-197

46. Smlier, S.W., 1985., "Problem Solving in the Domain of Qualitative Well Log Interpolation", The Second Conference on AI Application, Miami Beach,FL, U.S.A. pp. 55 ff

47. Sponemann, P. and L.Fuchs, 1986., "use of Artificial Intelligence for Real ime Control", in 48, pp. 415-422

48. Sriram, D. and R.Adey (eds.), 1986., **Applications of AI in Engineering Problems**, Vol II, Springer Verlag

49. Stedinger, J.R., 1980, "Fitting Lognormal Distribution to Hydrologic Data", Water Resource Research, Vol 16(3, pp.481-490

50. U.S.W.R.C., 1981., "Guidelines for Determining Flood Frequency", Bulletine 17B of the Hydrologic Committee, U.S. Govt. Printing Office, Washington, D.C.

51. Wilson, J.L. et al, 1986. "A Knowledge Based System for Hazard Site Evaluation", in 48, pp. 661-671

52. Yevjevich, V.,1972., **Probability and Statistics in Hydrology**, Water Resources Publication Colrado, U.S.A.

APPENDIX A

```
#include "clips.h"
run_clips(limit)
int *limit;
{
  run(*limit);
 }
c_advice()
{
 float num;
 char *name;
 num=rfloat(1);
 name=rstring(2);
 advice(&num,name);
 return(0);
}
extern int parameter();
usrfuncs()
{
 define_function("parameter",'i',parameter,"parameter");
 define_function("advice",'i',c_advice,"c_advice");
}
```

APPENDIX - B

VAX-VMS Linking Commands

1) copy all the CLIPS include files to your directory:

   $copy [{CLIPS master directory}]*.h [{your directory}]

2) create an object file from runclips.c:

   $cc runclips.c

3) compile the FORTRAN routine(s)

   $fortran {FORTRAN files}

4) link all the files together.

   $link/executable={exec name} {FORTRAN files}, **runclips, [{CLIPS** master directory}] **clipslib/library**

Note that one of the FORTRAN programs <u>must</u> be a main program. The CLIPS library should include the **loadc**, and **storec** functions.

APPENDIX C

ES Interface


```
C       *****************************************************
C       E X P E R T   S Y S T E M   I N T E R F A C E
C       *****************************************************
        SUBROUTINE PIHELP(A,SK,ST,V,AK,N)
        CHARACTER *50 FACT1,FACT2,FACT3,FACT4,FACT5,INT_FILE,FILE
        INTEGER C_FACT,C_FILE,K,CC_FACT,CCC_FACT,CCF_FACT,CCN_FACT
 1000   FORMAT(E12.5)
 2000   FORMAT(I5)
        IF(KKK.NE.1)CALL RESET_CLIPS
        CALL INIT_CLIPS
        FILE= 'FACHVES.CLP'
        CALL STOREC(FILE,C_FILE)
        K=LOAD_RULES(C_FILE)
        CALL RESET_CLIPS
        CALL DISPLAYFACTS()
        WRITE(INT_FILE,1000)A
        FACT1='mean '//INT_FILE(1:12)
        CALL STOREC(FACT1,C_FACT)
        CALL ASSERT(C_FACT)
        WRITE(INT_FILE,1000)SK
        FACT2='skew '//INT_FILE(1:12)
        CALL STOREC(FACT2,CC_FACT)
        CALL ASSERT(CC_FACT)
        WRITE(INT_FILE,1000)AK
        FACT3='kurt '//INT_FILE(1:12)
        CALL STOREC(FACT3, CCC_FACT)
        CALL ASSERT(CCC_FACT)
        WRITE(INT_FILE,1000)ST
        FACT4='stdev '//INT_FILE(1:12)
        CALL STOREC(FACT4, CCF_FACT)
        CALL ASSERT(CCF_FACT)
        WRITE(INT_FILE,2000)N
        FACT5='length of data '//INT_FILE(1:5)
        CALL STOREC(FACT5, CCN_FACT)
        CALL ASSERT(CCN_FACT)
        CALL DISPLAYFACTS()
        k=RUN_CLIPS(-1)
        CALL DISPLAYFACTS()
        TYPE *,'Now you are out of clips'
        RETURN
        END
```

28-Nov-1990 09:33:56 VAX FORTRAN V4.6-244
28-Nov-1990 09:33:27 $DISK2:[JK.MISHRA]FORCLIP.FOR;3

```
C       ******************************************************
C       ________________________
        SUBROUTINE STOREC(CHARS,CSTRING)
        INTEGER CSTRING (1)
        CHARACTER * (*) CHARS
        IWORD =1
        IBIT = 0
        K = LENGTH (CHARS)
        DO 100 I=1,K
        IVALUE=ICHAR(CHARS(I:I))
        CALL MVBITS (IVALUE,0,8,CSTRING(IWORD),IBIT)
        IBIT=IBIT+8
        IF(IBIT .LT. 32)GO TO 100
        IBIT=0
        IWORD=IWORD +1
  100   CONTINUE
        CALL MVBITS(0,0,8,CSTRING(IWORD),IBIT)
        RETURN
        END
```

28-Nov-1990 09:33:27 $DISK2:[JK.MISHRA]FORCLIP.FOR;35

```
C       __________________________________________
C       **********************
C       _____________________________________
        INTEGER FUNCTION LENGTH(STRING)
        CHARACTER * (*) STRING
        K=LEN(STRING)
        DO 100 I=K,1,-1
        IF(STRING(I:I) .NE.' ') GO TO 150
  100   CONTINUE
  150   CONTINUE
        LENGTH=I
        RETURN
        END
```

28-Nov-1990 09:33:27 $DISK2:[JK.MISHRA]FORCLIP.FOR;35

```
C       *************************************
        SUBROUTINE LOADC(CHARS,CSTRING)
        INTEGER CSTRING (1)
        CHARACTER * 1 CHAR
        CHARACTER * (*) CHARS
        IWORD=1
        IVALUE=0
        CHARS=' '
        IBIT=0
        K=LEN (CHARS)
        DO 100 I=1,K
        CALL MVBITS(CSTRING(IWORD),IBIT,8,IVALUE,0)
        IF(IVALUE .LT. 32)GO TO 200
        CHARS(I:I)=CHAR(IVALUE)
        IBIT=IBIT+8
        IF(IBIT.LT.32)GO TO 100
        IBIT=0
        IWORD=IWORD+1
  100   CONTINUE
  200   CONTINUE
        RETURN
        END
```

```
C       **********************************
        SUBROUTINE ADVICE(VALUE,C_NAME)
        INTEGER C_NAME
        REAL VALUE
        CHARACTER *80 NAME
        character *80 c_c_name
        common/cl_ret/v_value,c_c_name
        CALL LOADC(NAME,C_NAME)
        c_c_name=name
        v_value=value
c       TYPE *,'NAME=',NAME
c       TYPE *,'VALUE=',VALUE
        RETURN
        END
```

28-Nov-1990 09:33:56 VAX FORTRAN V4.6-244
28-Nov-1990 09:33:27 $DISK2:[JK.MISHRA]FORCLIP.FOR;3

APPENDIX D

Main Program

28-Nov-1990 09:33:56 VAX FORTRAN V4.6-244
28-Nov-1990 09:33:27 $DISK2:[JK.MISHRA]FORCLIP.FO

```
C       A TESTING PROGRAM
        REAL KURT
        CHARACTER *80 C_C_NAME
        COMMON/BLACK1/X(600),N,TITLE(80)
        COMMON/BLOCK1/NS,NPT,IALT
        COMMON/CL_RET/V_VALUE,C_C_NAME
        COMMON/BLOCK2/RX(600),NX,HEADING(80)
        DIMENSION TX(600)
        TYPE *,'**************************************************'
        TYPE *,'*      MASTERS THESIS OF O.P.MISHRA              **'
        TYPE *,'*       UNDER  THE  GUIDANCE    OF               **'
        TYPE *,'*      Dr.S.Ramaseshan and Dr.B.Dutta            **'
        TYPE *,'*           at I.I.T.KANPUR                      **'
        TYPE *,'**************************************************'
        TYPE *,'*    WISH   YOU    HAPPY COMPUTING               **'
        TYPE *,'*                W I T H                         **'
        TYPE *,'*             F A C H V E S                      **'
        TYPE *,'**************************************************'
        TYPE *,'* *--* FREQUENCY-ANALYSIS WITH EXPERT ADVICE *--*'
        TYPE *,'* ---------------------------------------------*'
11      TYPE *,'******************************************'
        TYPE *,'*      I N P U T    M O D U L E         *'
        TYPE *,'* -------------------------------------*'
        TYPE *,'Do you want to give input on the screen?'
        TYPE *,'If yes, Type 1 ,If no, Type 2'
        TYPE *,'    If you want to quit Type 3'
        READ(*,*)INP
        GO TO (111,19,9)INP
19      WRITE(*,*)'--The input file should be named as TEST.INP'
        WRITE(*,*)'--          It must have following :'
        TYPE *,'--      *   a TITLE with format 80A1 '
        TYPE *,'--             AND, unformated          '
        TYPE *,'--      *   NX--Number of data          '
        TYPE *,'--      *   NS--Number of seasons       '
        TYPE *,'--        (for annual series NS =1)     '
        TYPE *,'--      *   NPT--Number of points       '
        TYPE *,'--          NPT=NCLASS+1 where NCLASS is number of'
        TYPE *,'              classes in which data can be divided'
        TYPE *,'              for Chisquare test. The class frequency'
        TYPE *,'              should be equal to 5'
        TYPE *,'--          (It can be 7/11/13)         '
        TYPE *,'--      *   Observational hydrologic data RX(I)'
        TYPE *,'--      (Not more than 600;'
        TYPE *,'               seasons(in row)/years(in column)'
        TYPE *,'Have you got such an input file ? If yes,type 1'
        TYPE *,'        If you want to quit ,type 2'
        TYPE *,'        If you want to go to previous menu type 3'
        READ(*,*)IC
        GO TO (110,90,11)IC
110     CALL INPUT
        TITLE(80)=HEADING(80)
        REWIND 20
        GO TO 112
111     WRITE(*,*)'LENGTH OF DATA?'
        READ(*,*)NX
        WRITE(*,*)'DATA; X(I)='
        READ(*,*)(RX(I),I=1,NX)
```

```
        IALT=0
        WRITE(*,*)'GIVE NS, NPT'
        WRITE(*,*)'NS--No. of seasons,NPT-No of points'
        READ(*,*)NS,NPT
112     N=NX/NS
113     IF(NS.GT.1)THEN
        WRITE(*,*)'For which season do you want to do the analysis ?'
        WRITE(*,*)'Type the serial no of the season'
        READ(*,*)NC
        IF(NC.GT.NS)GO TO 180
        ELSE
        NC=1
        ENDIF
        LK=(NC-1)*N
        DO 12 IJK=LK+1,LK+N
12      X(IJK-LK)=RX(IJK)
        IF(NS.GT.1)WRITE(*,1111)NC
1111    FORMAT(4X,'THIS COMPUTATION IS FOR SEASON NO.',5I//)
9       TYPE *,'*-----------------------------------------------*'
        TYPE *,'*      MASTER-MENU                              *'
        TYPE *,'*-----------------------------------------------*'
        TYPE *,'*   1.PRELIMINARY ANALYSIS                      *'
        TYPE *,'*   2.CHOICE OF DISTRIBUTIONS                   *'
        TYPE *,'*   3.TESTS FOR GOODNESS OF FIT                 *'
        TYPE *,'*   4.PLOTTING OF RESULTS                       *'
        TYPE *,'*   5.QUIT                                      *'
        TYPE *,'*   6.HELP FROM EXPERT SYSTEM INTERFACE         *'
        IF(NS.GT.1)THEN
        TYPE *,'*   7.ANALYSIS FOR OTHER SEASONS                *'
        ENDIF
        TYPE *,'*-----------------------------------------------*'
        TYPE *,'*************************************************'
        IF(IES.NE.0)THEN
        TYPE *,'VALUE PASSED IS ',V_VALUE
        TYPE *,'NAME PASSED IS ',C_C_NAME
        ENDIF
        TYPE *,'Type the serial number of desired option'
        READ(*,*)MM
        IF(MM.EQ.7)GO TO 113
        GO TO (20,40,50,60,90,100)MM
20      TYPE *,'******************************************'
        TYPE *,'*   $$$ PRELIMINARY ANALYSIS $$$         *'
        TYPE *,'*----------------------------------------*'
        TYPE *,'*   1.STATISTICAL PARAMETERS             *'
        TYPE *,'*   2.FREQUENCY PLOT                     *'
        TYPE *,'*   3.C.D.F.PLOT                         *'
        TYPE *,'*   4.TRANSFORMATIONS                    *'
        TYPE *,'*   5.HELP FROM EXPERT SYSTEM INTERFACE  *'
        TYPE *,'*   6.TO MASTER MENU                     *'
        TYPE *,'*----------------------------------------*'
        TYPE *,'Type the serial number of desired option'
        READ(*,*)IPA
        GO TO (21,22,22,31,100,9)IPA
21      CALL PARAM(X,N,AMEAN,VAR,SKEW,KURT,STDEV)
        GO TO 20
22      GO TO 60
30      TYPE *,'******************************************'
```

```
          TYPE *,'**   $$ CHOICE OF DISTRIBUTION $$      **'
          TYPE *,'*----------------------------------------*'
          TYPE *,'*   1.NORMAL                        *'
          TYPE *,'*   2.LOG-NORMAL (2-PARAMETER)  *'
          TYPE *,'*   3.LOG-NORMAL (3-PARAMETER)  *'
          TYPE *,'*   4.PEARSON TYPE 3            *'
          TYPE *,'*   5.LOG-PEARSON TYPE-3        *'
          TYPE *,'*   6.EXTREME-VALUE TYPE-1      *'
          TYPE *,'*   7.EXTREME-VALUE TYPE-3      *'
          TYPE *,'*   8.FAP--PROGRAMS             *'
          TYPE *,'*   9.TO MASTER MENU            *'
          TYPE *,'*  10.TRANSFORMATIONS           *'
          TYPE *,'*  11.METHOD OF ANALYSIS        *'
          TYPE *,'*-----------------------------------------*'
          TYPE *,'If you want to transform the data**TYPE-10'
          TYPE *,'If not,then which distribution? type the serial number'
          READ(*,*)IOC
          GO TO (32,33,34,35,36,37,38,39,9,31,40)IOC
 32       CALL ND
          GO TO 50
 33       CALL LN2(METHOD)
          GO TO 50
 34       CALL LN3(METHOD)
          GO TO 50
 35       CALL PT3(METHOD)
          GO TO 50
 36       CALL LP3(METHOD)
          GO TO 50
 37       CALL T1E(METHOD)
          GO TO 50
 38       CALL T3E(METHOD)
          GO TO 50
 39       WRITE(*,*)'It fits normal distribution with or without
          1transformations'
          CALL FAP
          GO TO 50
 31       TYPE *,'*******************************************'
          TYPE *,'*   TRANSFORMATIONS                    *'
          TYPE *,'*   ----------------------------------*'
          TYPE *,'*   1.LOG-TRANSFORMATION               *'
          TYPE *,'*   2.PEARSON-TRANSFORMATION           *'
          TYPE *,'*   3.SQUARE-ROOT-TRANSFORMATION       *'
          TYPE *,'*   4.POWER-TRANSFORMATION             *'
          TYPE *,'*   5.CHOICE OF DISTRIBUTION           *'
          TYPE *,'*   6.TO MASTER MENU                   *'
          TYPE *,'*--------------------------------------*'
          TYPE *,'Type the serial number of desired option'
          READ(*,*)IT
          GO TO (311,312,313,314,30,9)IT
 311      CALL LOGTRAN(N,TX,X)
          GO TO 310
 312      CALL IPT(N,NS,AMEAN,STDEV,SKEW,X)
          DO 315 I=1,N
 315      TX(I)=X(I)
          GO TO 310
 313      CALL SQTRAN(N,X,TX)
          GO TO 310
```

```
  314   CALL TSPT(X,N,TX)
  310   WRITE(*,*)(TX(I),I=1,N)
        DO 316 I=1,N
  316   X(I)=TX(I)
        GO TO 30
 40     TYPE *,'***************************************************'
        TYPE *,'*  $$ METHOD OF ANALYSIS $$           *'
        TYPE *,'*-------------------------------------*'
        TYPE *,'*   1.METHOD OF MOMENTS               *'
        TYPE *,'*   2.METHOD OF MAXIMUM LIKELYHOOD    *'
        TYPE *,'*   3.METHOD OF LEAST-SQUARES         *'
        TYPE *,'*   4.TO MASTER MENU                  *'
        TYPE *,'*-------------------------------------*'
        TYPE *,'Which method do you like?'
        TYPE *,'Type the serial number of desired option'
        READ(*,*)MA
        GO TO (41,42,43,9)MA
 41     METHOD=1
        GO TO 30
 42     METHOD=2
        GO TO 30
 43     WRITE(*,*)'It is only available for Normal Distribution'
        WRITE(*,*)'For other cases pretransform the data into normal
        1          variables'
        WRITE(*,*)'Do you want to use it ? If yes type 1'
        READ(*,*)MLSQ
        IF(MLSQ.EQ.1)THEN
        CALL MLS
        ENDIF
        GO TO 40
 50     TYPE *,'***************************************************'
        TYPE *,'*  $$ TEST OF GOODNESS OF FIT $$                 **'
        TYPE *,'*-----------------------------------------------*'
        TYPE *,'*    1.CHI-SQUARE TEST                          *'
        TYPE *,'*    2.KOLMOGOROV-SMIRNOV TEST                  *'
        TYPE *,'*    3.TO PLOTTING RESULTS                      *'
        TYPE *,'*    4.COMPARISION OF STANDARD ERROR            *'
        TYPE *,'*      OF ESTIMATE FOR DIFFERENT                *'
        TYPE *,'*        DISTRIBUTIONS                          *'
        TYPE *,'*    5.TO MASTER MENU                           *'
        TYPE *,'*    6.HELP FROM EXPERT-SYSTEM INTERFACE        *'
        TYPE *,'*-----------------------------------------------*'
        TYPE *,'Type the serial number of desired option'
        READ(*,*)ITOG
        GO TO (51,52,53,54,9,100)ITOG
  51    IF(IT.EQ.0)GO TO 512
        IF(IT-4)511,511,512
  512   DO 513 I=1,N
  513   TX(I)=X(I)
  511   CALL CHISQRT(8,AMEAN,STDEV,TX,N,X)
        GO TO 60
   52   CALL KOLMG(X,N,D)
        GO TO 60
   53   TYPE *,'Are you satisfied? If yes type 1'
        READ(*,*)IS
        IF(IS.EQ.1)GO TO 60
        GO TO 9
```

```
 54     CALL SER
        GO TO 9
 60     TYPE *,'************************************************'
        TYPE *,'**  $$ PLOTTING RESULTS                      *'
        TYPE *,'*---------------------------------------*'
        TYPE *,'*   1.FREQUENCY PLOT                  *'
        TYPE *,'*   2.C.D.F.PLOT                      *'
        TYPE *,'*   3.TO MASTER MENU                  *'
        TYPE *,'*----------------------------------------*'
        TYPE *,'Type the serial number of desired option'
        READ(*,*)IPR
        GO TO (61,62,9)IPR
 61     TYPE *,'NOT INCORPORATED AS YET'
 62     TYPE *,'NOT INCORPORATED AS YET'
 100    IF(AMEAN.EQ.0.0)THEN
        TYPE *,'Expert system interface is only available'
        TYPE *,'    when statistical parameters are known'
        GO TO 20
        ENDIF
        TYPE *,'*********************************************'
        TYPE *,'*  $$ EXPERT SYSTEM INTERFACE $$         **'
        TYPE *,'*---------------------------------------*'
        TYPE *,'*   1.FOR PRELIMINARY ADVICE            *'
        TYPE *,'*   2.FOR SECONDARY   ADVICE            *'
        TYPE *,'*   3.FOR EXPERT ANALYSIS OF RESULTS    *'
        TYPE *,'*   7.TO MASTER MENU                    *'
        TYPE *,'*---------------------------------------*'
        TYPE *,'Which help do you want?'
        TYPE *,'Type the serial number of desired option'
        READ(*,*)IES
        IF(IES.EQ.1)GO TO 101
        GO TO (101,102,103,9)IES
        IF(IES.EQ.7)GO TO 9
 101    CALL PTHELP(AMEAN,SKEW,STDEV,VAR,KURT,N)
        TYPE *,'Value passed = ',v_value
        TYPE *,'Name passed is = ',c_c_name
        GO TO 9
 102    TYPE *,'NOT INCORPORATED AS YET'
 103    TYPE *,'NOT INCORPORATED AS YET'
        GO TO 9
 180    TYPE *,'There is no more seasons.'
        TYPE *,'*--------------------------*'
        TYPE *,'*   1.TO MASTER MENU      *'
        TYPE *,'*   2.TO INPUT MODULE     *'
        TYPE *,'*   3.TO QUIT             *'
        TYPE *,'*--------------------------*'
        TYPE *,'Type the serial number of desired option'
        READ(*,*)KL
        GO TO(9,11,90)KL
 90     STOP
        END
```

APPENDIX E

KBES

```
(deffacts a
(start))


(defrule rule1
(start)
=>
(fprintout t '       W E L C O M E                    'crlf)
(fprintout t '                                        'crlf)
(fprintout t '          TO                            'crlf)
(fprintout t '                                        'crlf)
(fprintout t '        F A C H V E S                   'crlf)
(fprintout t '                                        'crlf)
(fprintout t 'This  expert   system   is  designed  for  single
              peaked  continuous hydrologic variables   only'crlf)
(fprintout t '                                             'crlf)
(fprintout t 'IF YOU DO NOT UNDERSTANT THE QUESTION ; Type nk'crlf)
(fprintout t '                                              'crlf)
(fprintout t 'Type of data ---- type the first letter (e/a/s)
              Whether it is exceedence OR annual ?')
(assert(type of data =(read))))


(defrule rule2
(length of data ?n)
(type of data ?t)
=>
(if (< ?n 10)
then
(fprintout t 'No analysis is possible  ')
(assert(data is too small))
(assert(process continue no))
(bind ?num 1.0)
(assert(suggestion ?num))
(assert(esad NA))
(assert(stop))
else
(if (< ?n 25)
then
(fprintout t 'Results are suspected ;Do you want to continue ? ')
(assert(process continue =(read)))
(assert(insufficient data))
(bind ?num 2.0)
(assert(suggestion ?num))
(assert(esad RS))
else
(assert(process continue yes))
(assert(data is large enough)))))
```

```
(defrule rule3
(declare(salience 10))
(process continue ?pc)
(test (eq ?pc yes))
=>
(fprintout t "Do you think that outliers are present?  ")
(assert(outliers are =(read))))

(defrule rule4
(outliers are ?om)
(or(data is large enough)
   (insufficient data))
(process continue ?pc)
(test (eq ?pc yes))
=>
(if (eq ?om yes)
then
(assert(outliers yes))
else
(if (eq ?om nk)
then
(fprintout t "In absence of information outliers are
              assumed not to be present"crlf)
(assert(outliers no)))))


(defrule rule5
(type of data ?t)
(or(data is large enough)
   (insufficient data))
(process continue ?pc)
(test (eq ?pc yes))
(outliers ?o)
=>
(if (eq yes ?o)
then
(assert(outliers present))
else
(if (eq ?t e)
then
(fprintout t "Negative exponential distribution is well suited"crlf)
(bind ?num 4.0)
(assert(suggestion ?num))
(assert(esad NED))
else
(fprintout t "Here is the list of steps_____**
              1.**DO THE PRELIMINARY ANALYSIS
              2.**FIT A SUITABLE DISTRIBUTION
                (GENERALLY A 2 PARAMETER DISTRIBUTION
              3.**TEST FOR GOODNESS OF FIT
              4.**COMPARE THE RESULTS"crlf))))
```

```
(defrule rule6
(process continue ?pc)
(outliers present)
(type of data ?t)
(outliers ?o)
(test (eq ?o yes))
(test (eq ?pc yes))
=>
(if (eq ?t e)
then
(fprintout t 'You should use method of least squares for
              parameter estimation AND also use negative
              exponential distribution'crlf)
(assert(stop))
(bind ?num 28.0)
(assert(suggestion ?num))
(assert(esad MLSNED))
else
(fprintout t 'You should use method of least squares for
               parameter estimation AND also use normal
               distribution ;You should transform the data
               essentially to normal distribution using
               suitable transformations  'crlf)
(assert(stop))
(bind ?num 4.0)
(assert(suggestion ?num))
(assert(esad MLSND))))


(defrule rule7
(process continue ?pc)
(test (eq ?pc yes))
(or(data is large enough)
   (insufficient data))
(not(stop))
(kurt ?k)
(mean ?am)
(skew ?sk)
(stdev ?st)
(test(&&(&&(> ?sk -0.5)
           (< ?sk 0.5))
     (> 0.4(/ ?st ?am))))
=>
(if (< ?sk 0)
then
(if (< ?st(/ ?am 3.0))
then
(fprintout t 'NORMAL DISTRIBUTION CAN BE APPROXIMATED'crlf)
(bind ?num 11.0)
(assert(suggestion ?num))
(assert(esad ND))
else
(if (> ?k 1.0)
then
(if (< ?k 2.25)
then
```

```
(fprintout t "UNIFORM DISTRIBUTION IS WELL SUITED"crlf)
(bind ?num 5.0)
(assert(suggestion ?num))
(assert(esad UD))
else
(if (&&(> ?k 3.5)
       (< ?k 10.0))
then
(fprintout t "t- DISTRIBUTION IS WELL SUITED"crlf)
(bind ?num 7.0)
(assert(suggestion ?num))
(assert(esad tD))
else
(if (&&(< ?k 3.5)
       (> ?k 2.25))
then
(fprintout t "NORMAL DISTRIBUTION IS WELL SUITED"crlf)
(bind ?num 6.0)
(assert(suggestion ?num))
(assert(esad ND)))))))))


(defrule rule8
(or(data is large enough)
   (insufficient data))
(process continue ?pc)
(test (eq ?pc yes))
(kurt ?k)
(mean ?am)
(skew ?sk)
(test(||(<= ?sk -0.5)
        (>= ?sk 0.5)))
=>
(if (<= ?sk -0.5)
then
(assert(negative skew))
else
(assert(positive skew))))


(defrule rule9
(or(data is large enough)
   (insufficient data))
(process continue ?pc)
(test (eq ?pc yes))
(positive skew)
(kurt ?k)
(skew ?sk)
(test(< ?k (+ (* 1.25 ?sk) 1)))
=>
(fprintout t "No distribution will be suited"crlf)
(fprintout t " Check for data errors or multiple distribution"crlf)
(bind ?num 20.0)
(assert(suggestion ?num))
(assert(esad NoD)))
```

```
(defrule rule10
(positive skew)
(process continue ?pc)
(or(data is large enough)
   (insufficient data))
(test (eq ?pc yes))
(kurt ?k)
(skew ?sk)
(test(&&(< ?k (+ (* 2.25 ?sk) 3))
        (> ?k (+ (* 1.25 ?sk) 3))))
=>
(fprintout t 'Lognormal 2 parameter distribution is well suited'crlf)
(bind ?num 21.0)
(assert(suggestion ?num))
(assert(esad LN2)))


(defrule rule11
(positive skew)
(not(stop))
=>
(call(parameter))
(fprintout t 'From the screen type standard deviation of
              logtransformed data ?')
(assert(stdev of logtran data =(read))))


(defrule rule12
(positive skew)
(stdev of logtran data ?ltst)
=>
(bind ?lcv (sqrt (** 2.278 (- (** ?ltst 2.0) 1.0))))
(bind ?lest (* 3.0 ?lcv))
(bind ?lfst (** ?lcv 3.0))
(bind ?lsk (+ ?lest ?lfst))
(assert(logtran skew is ?lsk)))


(defrule rule13
(positive skew)
(logtran skew is ?lsk)
(coeff of var ?lcv)
(test(&&(&&(> ?lsk -0.5)
           (< ?lsk 0.5))
     (> 0.4 ?lcv)))
=>
(assert(lognormal2))
(fprintout t ' Following transformations can be done
              1.Log transformation
              2.N_root transformation
              OR; You can use 2 parameter lognormal distribution'crl:
(bind ?num 13.0)
(assert(suggestion ?num))
(assert(esad LN2)))
```

```
(defrule rule17
(positive skew)
(kurt ?k)
(not(stop))
(length of data ?n)
(test(&&(> ?k (+ 2.2 (/ 6.0 (sqrt ?n))))
       (< ?k (+ 3.8 (/ 6.0 (sqrt ?n))))))
=>
(fprintout t 'Gamma distribution is well suited'crlf)
(fprintout t 'Method of maximum liklihood is recommended'crlf)
(bind ?num 10.0)
(assert(suggestion ?num))
(assert(esad GD)))


(defrule rule18
(negative skew)
(skew ?sk)
(mean ?am)
=>
(fprintout t 'YOU CAN DO FOLLOWING TRANSFORMATIONS
              1.INVERSE-PEARSON TRANSFORMATION
              2.n-ROOT TRANSFORMATION
             OR;YOu can fit either pearson type 3
             OR;YOu can fit logpearson type 3
             OR; You can't fit lognormal 3 distribution'crlf)
(bind ?num 12.0)
(assert(suggestion ?num))
(assert(esad PT3,LP3,LN3,IPTRAN)))

(defrule rule19
(declare (salience -100))
(type of data ?t)
(process continue ?pc)
(test (eq ?pc yes))
=>
(if (eq ?t s)
then
(fprintout t '1.You can fit one distribution for all seasons;OR
              2.You can fit different distributions for different
                seasons;OR
              3.You can fit all available distributions for all seasons'crlf)
(assert(stop))
else
(if (eq ?t a)
then
(fprintout t '1.You can fit either one OR all available distributions'crlf)
(assert(stop)))))


(defrule rule20
(declare (salience -1000))
(stop)
=>
(fprintout t 'PROCESS IS BEING STOPPED;
              RESET THE PROCESS IF YOU WANT TO CONTINUE
              WITH OTHER SET OF DATA'crlf)
(fprintout t 'Do you want any help ?  ')
(assert(help =(read))))
```

```
(defrule rule21
(declare (salience -990))
(suggestion ?num)
(esad ?string)
=>
(advice ?num ?string)
(assert(stop)))

(defrule rule22
(stop)
(help ?hl)
=>
(if (eq no ?hl)
then
(fprintout t 'Suggestions passed should be taken in sequencial order'crlf)
(halt)))

(defrule rule24
(help ?hl)
=>
(if (eq yes ?hl)
then
(fprintout t 'I can help you in following ways--***
              1.Different types of distributions
              2.Different types of transformations
              3.Different types of goodness of fit tests
              4.References
              6.Reasoning for any concluding remarks'crlf)
(fprintout t 'MAY I HELP YOU?  ')
(assert(help required =(read)))))

(defrule rule25
(help ?hl)
(help required ?hp)
=>
(if (eq ?hp yes)
then
(fprintout t 'Type the serial number'crlf)
(fprintout t 'Which help do you need?   ')
(assert(help needed =(read)))
else
(fprintout t 'Then why did you disturb me????'crlf)))

(defrule rule26
(help needed ?hn)
(help ?hl)
=>
(if (= ?hn 1)
then
(fprintout t 'The following types of distributions are normally
 used in hydrology ----**----***
     **.NORMAL DISTRIBUTION
     **.LOG-NORMAL 2 DISTRIBUTION
     **.LOG-NORMAL 3 PARAMETER DISTRIBUTION
     **.PEARSON TYPE 3 DISTRIBUTION
     **.LOG-PEARSON TYPE 3 DISTRIBUTION
     **.TYPE 1 EXTREMAL DISTRIBUTION
     **.TYPE 3 EXTREMAL DISTRIBUTION'crlf)
else
```

```
(if (= ?hn 2)
then
(fprintout t 'The following transformations are generally
used in HYDROLOGY----****_____
        **.LOG
        **.n-TH ROOT
        **. INVERSE-PEARSON
        **.TWO STEP POWER'crlf)

else
(if (= ?hn 3)
then
(fprintout t 'Two tests can  presenty  be done
We are extremely sorry for these limitations
Anyway,they are sufficient for hydrologic data analysis
     *--*.CHI-SQUARE TEST(a parametric test)
     *--*.KOLMOGROV-SMIRNOV TEST(a non-parametric test)'crlf)
else
(fprintout t 'I can't help much.See the following references
       1.Applied Hydrology BY V.T.Chow & Maidment
       2.Handbook of Hydrology BY V.T Chow
       3.Frequency and Risk Analysis BY G.W.Kite
       4.Statistical Methods In Hydrology by Haan
       5.Hydrology by Raudkivi
       6.Probability and Statistics in Hydrology bY  Y.Yevjevich
       7.Hydrologic Frequency Modeling,edited by V.P.Singh'crlf)))))
```